AD-A040 773 AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TEX F/G 5/9
PROCEEDINGS, ANNUAL CONFERENCE OF THE MILITARY TESTING ASSOCIAT--ETC(U)
NOV 73

UNCLASSIFIED NL

1 OF 9
ADA
040773



AD A040773

PROCEEDINGS

(1)

B.S.

15th ANNUAL CONFERENCE of the MILITARY TESTING ASSOCIATION



COPY AVAILABLE TO DDC DOES NOT PERMIT FULLY LEGIBLE PRODUCTION

Coordinated By:
AIR FORCE HUMAN RESOURCES LABORATORY
AIR FORCE SYSTEMS COMMAND
Brooks Air Force Base, Texas

DDC
RECEIVED
JUN 21 1977
D

Approved for public release; distribution unlimited.

El Tropicano Motor Hotel
San Antonio, Texas
28 October - 2 November 1973

AD No.
DDC FILE COPY

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. AVAIL. and/or SPECIAL
A 23

DTC

(6) PROCEEDINGS

(1)

~~15th~~ Annual Conference

of the

Military Testing Association (15th)

Held at

(11) 2 Nov 73

(12) 831p.

Approved for public release; distribution unlimited.

Coordinated By:

AIR FORCE HUMAN RESOURCES LABORATORY
AIR FORCE SYSTEMS COMMAND
Brooks Air Force Base, Texas

El Tropicano Motor Hotel
San Antonio, Texas
28 October - 2 November 1973

DDC
RECEIVED
JUN 21 1977
D

AD A040415

AB

FOREWORD

The 15th Annual Conference of the Military Testing Association was held at the El Tropicano Motor Hotel, San Antonio, Texas, 28 October - 2 November 1973.

The papers presented at the 15th Annual Conference of the MTA reflect a diversity of subject matter. These presentations from representatives of both the military and civilian community reflect the opinions of the authors and are not to be construed as official or any way representative of the views and policies of the United States Armed Services or the armed services of the foreign countries participating in this meeting.

PRECEDING PAGE BLANK-NOT FILMED

- OFFICIAL PROGRAM -

15th Annual Conference
MILITARY TESTING ASSOCIATION

28 October through 2 November 1973

	Sunday, October 28
El Tropicano Lobby 1300 - 1700	Registration
	Monday, October 29
El Tropicano Lobby 0800 - 1600	Registration
Hemisfair Room Ground Floor	Session Chairman Col Theodore B. Aldrich Air Force Human Resources Laboratory
0900	Conference called to order
0900 - 0915	Greetings by AFHRL Commander Col Harold E. Fischer Air Force Human Resources Laboratory
0915 - 0945	Keynote Address Lt Gen John W. Roberts Deputy Chief of Staff Personnel Headquarters, United States Air Force
0945 - 1010	Introductory Comments Col Theodore B. Aldrich Air Force Human Resources Laboratory
1010 - 1015	Administrative Announcements
1015 - 1030	Coffee Break

PRECEDING PAGE BLANK-NOT FILMED

1030 - 1200	Symposium *Early-service screening for military performance and behavior* Chairman Dr. Ralph R. Canter Office of the Assistant Secretary of Defense (M&RA) Participants Maj James Taylor United States Air Force Comdr James Murphy United States Navy Lt Col Todd Graham United States Army Dr. D. Brian Murphy Systems Development Corporation Dr. Robert G. Smith Human Resources Research Organization Discussant Dr. Norman J. Kerr Naval Technical Training Command
1200 - 1330	Lunch
Hemisfair Room Ground Floor	Session Chairman Mr. Harald E. Jensen Air Force Human Resources Laboratory
1330 - 1430	Symposium *Cooperative aptitude research programs conducted through the Armed Forces Vocational Testing Group* Chairman Dr. Lonnie D. Valentine Air Force Human Resources Laboratory Participants Maj Verna S. Kellogg III Armed Forces Vocational Testing Group Capt Melvin T. Gambrell, Jr. Armed Forces Vocational Testing Group Dr. Harry D. Wilfong Armed Forces Vocational Testing Group

1430 - 1450	*Development of the Armed Services Vocational Aptitude Battery* Dr. M. A. Fischl United States Army Research Institute for the Behavioral and Social Sciences
1450 - 1505	Coffee Break
	Session Chairman Capt James A. Hoskins Air Force Human Resources Laboratory
1505 - 1535	*New developments in defense language proficiency testing* Drs. C. D. Leatherman and A. Al-Haik United States Army, Defense Language Institute Presented by: Mr. Sydney Sako
1535 - 1550	*Enlisted selection and classification testing in the U. S. Army* Dr. Milton H. Maier United States Army Research Institute for the Behavioral and Social Sciences Presented by: Dr. M. A. Fischl
1550 - 1610	*The development of English Comprehension Level screening tests for foreign students* Mr. Cortez Parks United States Army, Defense Language Institute

END OF MONDAY SESSIONS

Monterrey Suite Room 909	
1730 - 1830	MTA President's welcoming and social hour
	Tuesday, October 30
Acapulco Room Room 902	
0800 - 1600	Registration

Hemisfair Room Ground Floor	Session Chairman Capt James A. Hoskins Air Force Human Resources Laboratory
0845 - 0900	Call to order
0900 - 0920	*An experimental, multimedia career development course for new mental standards airmen* Mr. George P. Scharf USAF School of Applied Aerospace Sciences
0920 - 0945	*Optimal utilization of on-the-job training and technical training school* Capt Alan D. Dunham National Security Agency
0945 - 1000	Coffee Break
	Session Chairman Mr. Manuel Pina Air Force Human Resources Laboratory
1000 - 1155	Symposium *Translation of training research into training action - a missing link* Chairman Dr. G. Douglas Mayo Naval Technical Training Command Participants *From the viewpoint of a training manager* Mr. Walter E. McDowell Army Training and Doctrine Command *From the viewpoint of a training researcher* Dr. Norman J. Kerr Naval Technical Training Command Discussants Lt Col Donald F. Mead Air Training Command Dr. Earl I. Jones Navy Personnel Research & Development Center

	Reaction Panel Comdr Bruce Cormack Canadian Armed Forces Mr. William B. Lecznar Air Force Human Resources Laboratory Mr. Ralph Canter Office of the Assistant Secretary of Defense (M&RA)
1155 - 1200	Administrative Announcements
1200 - 1330	Lunch
Monterrey Suite Room 909	
1330 - 1500	Meeting of the Steering Committee
Hemisfair Room Ground Floor	Session Chairman Mr. William J. Stacy Air Force Human Resources Laboratory
1330 - 1340	Administrative Announcements
1340 - 1400	*Test feedback and training objectives* LtJG Carroll H. Greene United States Coast Guard Training Center
1400 - 1420	*"PROBE" - An automated testing system for the evaluation of ability and/or willingness to profit from a training situation* Dr. David B. Vinson Assessment Systems, Incorporated
1420 - 1445	*Prospective chief petty officer leadership/management seminar* Comdr G. C. Hinson United States Coast Guard Training Center
1445 - 1510	*Transcendental Meditation (TM): A threat or technique in future personnel evaluation* Dr. R. O. Waldkoetter Consultant, Education and Personnel Systems
1510 - 1525	Coffee Break

	Session Chairman Capt Nicholas C. Varney Air Force Human Resources Laboratory
1525 - 1545	*Implications of carrel instruction on military testing* Dr. Ronald W. Spangenberg Air Force Human Resources Laboratory Presented by: Dr. Roger J. Pennell
1545 - 1600	*Assessment of complex psychomotor coordination learning and performance during 130 days of intensive testing* Dr. Randall M. Chambers Georgia Institute of Technology Dr. Rayford T. Saucer Veterans Administration Hospital
1600 - 1630	*Proficiency measurement in flight simulators* Dr. Edwin Cohen The Singer Company, Simulation Products Division

END OF TUESDAY SESSIONS

	Wednesday, October 31
Hemisfair Room Ground Floor	Session Chairman Mr. Andrew T. Garza Air Force Human Resources Laboratory
0830	Call to order
0830 - 0850	*The personnel and training evaluation program: A working program for improving the efficiency and effectiveness of fleet ballistic missile weapons system training-Part I - scope and achievement* Lt Comdr C. L. Walker Naval Guided Missiles School
0850 - 0920	*The personnel and training evaluation program: A working program for improving the efficiency and effectiveness of fleet ballistic missile weapons system training-Part II - program development* Mr. Frank B. Braun Data Design Laboratories

0920 - 0935	*Inventory and assessment at the United States Army War College* Lt Col Quay C. Snyder United States Army War College
0935 - 0950	*Subordinate ratings: Why not?* Drs. W. H. Githens and R. S. Elster United States Naval Postgraduate School
0950 - 1005	*The Keesler Study-electronic technicians four year evaluation of three types of training* Dr. Virginia Zachert Medical College of Georgia
1005 - 1020	Coffee Break
	Session Chairman Mr. Kenneth G. Koym Air Force Human Resources Laboratory
1020 - 1035	*Process versus product measures in performance testing* Mr. William C. Osborn Human Resources Research Organization
1035 - 1050	*The Self Evaluation Technique [SET] studies* Dr. John J. Holden United States Army Ordnance Center and School
1050 - 1120	*A modular approach to proficiency testing* Dr. Robert W. Stephenson, Mr. Warren P. Davis, and Mr. Harry I. Hadley American Institutes for Research Mrs. Bertha H. Cory United States Army Research Institute for the Behavioral and Social Sciences
1120 - 1200	*The growing demand for human performance testing* Mr. J. E. Gerber, Jr. United States Army Infantry School Presented by: Capt John P. Otjen
1200 - 1330	Lunch
Hemisfair Room Ground Floor	Session Chairman Lt James W. Abellera Air Force Human Resources Laboratory

1330 - 1350	*Criterion-referenced performance testing in combat arms skills* Mr. John F. Hayes URS/Matrix Company
1350 - 1420	*The Air Force wife: Her knowledge of, and attitudes toward, the Air Force* Drs. John A. Belt and Arthur B. Sweney Wichita State University
1420 - 1435	*Cadet retention at the Royal Military College of Canada as a function of personality, leadership performance, and biographical variables* Lt Col G. J. Carpenter Royal Military College of Canada
1435 - 1450	*Motivation of enlisted reservists in a non-draft atmosphere* Lt Knight D. Cheney United States Coast Guard Reserve
1450 - 1505	*Encounter groups as a means for effecting attitudinal change* Dr. Peter F. Newton National Security Agency
1505 - 1520	Coffee Break
	Session Chairman Mr. James M. Wilbourn Air Force Human Resources Laboratory
1520 - 1540	*An evaluation of diagnostic vocabulary testing in Air Force technical training* Lt William P. Mockovak Air Force Human Resources Laboratory
1540 - 1600	*Uniformed physicians without patients: A squandering of scarce resources?* Dr. Gary B. Brumback American Institutes for Research
1600 - 1645	*An experimental implementation of computer assisted admissible probability testing* Mr. W. L. Sibley RAND Corporation

	END OF WEDNESDAY SESSIONS
Monterrey Suite Room 909 1730 - 1830	MTA President's social hour
	Thursday, November 1
Hemisfair Room Ground Floor	Session Chairman Lt Barry P. McFarland Air Force Human Resources Laboratory
0830	Call to order
0830 - 0900	*CHAROSEL: A computer program which selects qualitative predictors for qualitative criterion prediction problems* Dr. William J. Moonan Navy Personnel Research and Development Center
0900 - 0915	*A factor analytic approach to the criterion problem* Dr. Roger Pennell Air Force Human Resources Laboratory
0915 - 1005	*Applications of predictor ordering and selection by Bayesian-decision technique* Mr. S. E. Bowser Navy Personnel Research and Development Center
1005 - 1020	Coffee Break
	Session Chairman Mr. William J. Phalen Air Force Human Resources Laboratory
1020 - 1045	*How admissible probability testing affects and is affected by the larger system of incentives* Dr. Thomas A. Brown RAND Corporation

1045 - 1100	*Growing demands on human resources: A view from industry* Dr. Thomas C. Tuttle Westinghouse Behavioral Safety Center
1100 - 1130	*Review of Air Force job satisfaction and career development research* Mr. R. Bruce Gould Air Force Human Resources Laboratory
1130 - 1145	*A methodology for determining job satisfaction in the U. S. Navy* Dr. Lawrence A. Goldman Naval Occupational Task Analysis Program
1145 - 1200	*Interest surveys and job placement in the Armed Services* Dr. William J. Figel Science Research Associates
1200 - 1330	Lunch
Hemisfair Room Ground Floor	Session Chairman Mr. C. Amos Johnson Air Force Human Resources Laboratory
1330 - 1355	*Job description in the Armed Forces as a basis for analysis and evaluation* Flottillenadmiral Guenter Fiebig Armed Forces Staff German Federal Ministry of Defense Presented by: Lt Col H. E. Seuberlich
1355 - 1420	*Possibilities and limitations of job analysis and job evaluation in the Armed Forces* Lt Col H. E. Seuberlich Armed Forces Staff German Federal Ministry of Defense
1420 - 1440	*Occupational analysis in the Royal Australian Air Force* Capt Wayne S. Sellman United States Air Force Sqn Ldr John W. K. Fugill Royal Australian Air Force Presented by: Sqn Ldr William J. Watson Air Force Human Resources Laboratory

1440 - 1500	*Relative aptitude requirements research* Dr. Raymond E. Christal and Sqn Ldr John W. K. Fugill, RAAF Air Force Human Resources Laboratory
1500 - 1515	Coffee Break
	Session Chairman Mr. Wayne S. Archer Air Force Human Resources Laboratory
1515 - 1540	*Collection of performance data at the job task level* Mr. Clifford P. Hahn American Institutes for Research
1540 - 1610	*Cognitive complexity: Its association with selection for military leadership roles* Sqn Ldr Brian N. Purry Royal Air Force
1610 - 1625	*New concepts for the measurement of attitudes and motives* Dr. Robert G. Smith, Jr. Human Resources Research Organization
1625 - 1630	Administrative Announcements

END OF THURSDAY SESSIONS

Terrace Ballroom 1900 - 2000	Social Hour - "pay-as-you-go" bar
2000 - 2200	Banquet - Invited Speaker Brig Gen Conrad S. Allman Commander, USAF Recruiting Service
	Friday, November 2
Hemisfair Room Ground Floor	Session Chairman Mr. William E. Alley Air Force Human Resources Laboratory
0845	Call to order

0845 - 0900	Administrative Announcements
0900 - 0920	*Administration of multiple-choice tests to non-readers via tape recorder: Case studies* Dr. Joseph L. Boyd, Jr. Educational Testing Service
0920 - 0940	*Racial differences in AFQT, AQE, and WAIS scores of Air Force basic trainees* Capt Stephen B. Knouse, Sgt David F. McGrevy, and Sgt Ronnie A. Thompson Air Force Human Resources Laboratory
0940 - 1005	*Does the USAF Officer Biographical Inventory portion of the AFOQT inadvertently measure authoritarian personality?* Lt Col Charles W. Haney Air Force Reserve Officer Training Corps
1005 - 1020	Coffee Break
	Session Chairman Mr. Larry T. Looper Air Force Human Resources Laboratory
1020 - 1035	*Human relations evaluation* Ms. A. M. Kelleher United States Coast Guard Training Center
1035 - 1100	*Centralized selection system for Air Force pilots* Lt Col Melvin S. Majesty USAF School of Military Sciences, Officer
1100 - 1130	*United States Army pilot assessment center project* Col Wallace F. Veaudry United States Army Infantry School
1130 - 1150	*Assessment centers: Validation of their use for selection, training, and career development in a military setting* Dr. Kay H. Smith United States Army Research Institute for the Behavioral and Social Sciences

1150 - 1210	*A proposed spatial orientation/disorientation flight training concept* Mr. Patrick J. Dowd USAF School of Aerospace Medicine
1210 - 1230	Concluding Comments

END OF CONFERENCE

TABLE OF CONTENTS

			Page
I.	Keynote Address	Lt General John W. Roberts	1
II.	Symposium synopsis: *Cooperative aptitude research programs conducted through the Armed Forces Vocational Testing Group*	Dr. L.D. Valentine	12
		Major Verna S. Kellogg III	12
		Capt Melvin T. Gambrell, Jr.	14
		Dr. Harry D. Wilfong	16
III.	*Development of the Armed Services Vocational Aptitude Battery*	Dr. M.A. Fischl	20
IV.	*New developments in defense language proficiency testing*	Dr. C.D. Leatherman	
		Dr. A. Al-Haik	27
V.	*Enlisted selection and classification testing in the U. S. Army*	Dr. Milton H. Maier	34
VI.	*The development of English Comprehension Level screening tests for foreign students*	Mr. Cortez Parks	45
VII.	*An experimental, multimedia career development course for new mental standards airmen*	Mr. George P. Sharf	52
VIII.	*Optimal utilization of on-the-job training and technical training school*	Capt Alan D. Dunham	60
IX.	Symposium: *Translation of training research into training action - a missing link*	Dr. G. Douglas Mayo	92
		Mr. Walter E. McDowell	96
		Dr. Norman J. Kerr	101
		Lt Col Donald F. Mead	108
		Dr. Earl I. Jones	114
X.	*Test feedback and training objectives*	Lt JG Carroll H. Greene	118

PRECEDING PAGE BLANK-NOT FILMED

			Page
XI.	*"PROBE" - An automated testing system for the evaluation of ability and/or willingness to profit from a training situation.*	Dr. David B. Vinson	129
XII.	*Prospective chief petty officer leadership/management seminar*	Comdr G. C. Hinson	141
XIII.	*Transcendental meditation (TM): A threat or technique in future personnel evaluation.*	Dr. R. O. Waldkoetter	147
XIV.	*Implications of carrel instruction on military testing.*	Dr. R. W. Spangenberg	154
XV.	*Proficiency measurement in flight simulators*	Dr. Edwin Cohen	170
XVI.	*The personnel and training evaluation program: A working program for improving the efficiency and effectiveness of fleet ballistic missile weapons system training - Part I: Scope and Achievement*	Lt Comdr C. L. Walker	178
XVII.	*The personnel and training evaluation program: A working program for improving the efficiency and effectiveness of fleet ballistic missile weapons system training - Part II: Program development.*	Mr. Frank B. Braun	191
XVIII.	*Inventory and assessment at the United States Army War College.*	Lt Col Quay C. Snyder	207
XIX.	*Subordinate ratings: Why not?*	Dr. W. H. Githens Dr. R. S. Elster	220
XX.	*The Keesler Study - electronic technicians four year evaluation of three types of training.*	Dr. Virginia Zachert	234
XXI.	*Process versus product measures in performance testing.*	Mr. William C. Osborn	249

			Page
XXII.	Brief of four research studies using the Self Evaluation Technique (SET studies)	Dr. John J. Holden	257
XXIII.	A modular approach to proficiency testing	Dr. Robert W. Stephenson Mr. Warren P. Davis Mr. Harry I. Hadley Mrs. Bertha H. Cory	269
XXIV.	The growing demand for human performance testing	Mr. J. E. Gerber, Jr.	297
XXV.	Criterion referenced performance testing in combat arms skills	Mr. John F. Hayes	358
XXVI.	The Air Force wife: Her knowledge of, and attitudes toward, the Air Force	Dr. John A. Belt Dr. Arthur B. Sweney	373
XXVII.	Cadet retention at the Royal Military College of Canada as a function of personality, leadership performance, and biographical variables	Lt Col G. J. Carpenter	395
XXVIII.	Motivation of enlisted reservists in a non-draft atmosphere	Lt Knight D. Cheney	411
XXIX.	An evaluation of diagnostic vocabulary testing in Air Force technical training	Lt William P. Mockovak	422
XXX.	Uniformed physicians without patients: A squandering of scarce resources?	Dr. Gary B. Brumback	441
XXXI.	An experimental implementation of computer assisted admissible probability testing	Mr. W. L. Sibley	460
XXXII.	CHAROSEL: A computer program which selects qualitative predictors for qualitative criterion prediction problems	Dr. William J. Moonan	483

		Page
XXXIII.	A factor analytic approach to the criterion problem. . . . . .Dr. Roger Pennell	508
XXXIV.	Applications of predictor ordering and selection by a Bayesian-decision technique. . . . . Mr. S. E. Bowser	525
XXXV.	Growing demands on human resources: A view from industry. . . . . Dr. Thomas C. Tuttle	540
XXXVI.	Review of Air Force job satisfaction and career development research . . . . . Mr. R. Bruce Gould	560
XXXVII.	A methodology for determining job satisfaction in the U. S. Navy . . . . . .Dr. Lawrence A. Goldman	573
XXXVIII.	Interest Surveys and job placement in the Armed Services . . . . . Dr. William J. Figel	585
XXXIX.	Job description in the Armed Forces as a basis for analysis and evaluation . . . . . Flottillenadmiral Guenter Fiebig	596
XL.	Possibilities and limitation of job analysis and job evaluation in Armed Forces . . . . . .Lt Col H. E. Seuberlich	608
XLI.	Occupational analysis in the Royal Australian Air Force . . . . . .Capt Wayne S. Sellman Sqn Ldr John W. K. Fugill	622
XLII.	Collection of performance data at the job task level . . . . . Mr. Clifford P. Hahn	643
XLIII.	Cognitive complexity: Its association with selection for military leadership roles . . . . . Sqn Ldr Brian N. Purry	659
XLIV.	New concepts for the measurement of attitudes and motives . . . . . Dr. Robert G. Smith, Jr.	674

			Page
XLV.	*Administration of muliple-choice tests to non-readers via tape-recorder: Case studies*	Dr. Joseph L. Boyd, Jr.	689
XLVI.	*Racial differences in AFQT, AQE, and WAIS scores of Air Force basic trainees*	Capt Stephen B. Knouse Sgt David F. McGrevy Sgt Ronnie A. Thompson	694
XLVII.	*Does the USAF Officer Biographical Inventory portion of the AFOQT inadvertently measure authoritarian personality?*	Lt Col Charles W. Haney	709
XLVIII.	*Human relations evaluation*	Ms. A. M. Kelleher	726
XLIX.	*Centralized selection system for Air Force pilots*	Lt Col Melvin S. Majesty	731
L.	*U. S. Army pilot assessment center project*	Col Wallace F. Veaudry	740
LI.	*Assessment centers: Validation of their use for selection, training, and career development in a military setting*	Dr. Kay H. Smith	750
LII.	*A proposed spatial orientation/ disorientation flight training concept*	Mr. Patrick J. Dowd	757
LIII.	Minutes of the 1973 MTA Steering Committee Meeting		789
LIV.	Citation of the Capt Harry J. Greer Award for 1973		793
LV.	By-Laws of the Military Testing Association		795
LVI.	List of conferees		801

KEYNOTE ADDRESS

Lt Gen John W. Roberts
Deputy Chief of Staff Personnel
Headquarters, United States Air Force

It's traditional for a speaker, particularly a keynote speaker, to open his remarks with the statement that he's pleased and gratified to be wherever he is. I can do this with a considerable measure of sincerity, since being here gives me an opportunity to pass on some thoughts to a group that holds the key to better use of human resources in these days of growing demands.

In the last few days, I've supplemented my knowledge about this audience with some homework, including a review of the minutes of your past meetings. Your achievements, both as individuals and as an association, are certainly more than impressive.

Although I do not speak to you as a psychologist or as a measurement specialist, I feel a definite kinship with many in this group. I have worked very closely with people in the Human Resources Laboratory for the past three years. As Director of Personnel Plans for the Air Force, and now as the Deputy Chief of Staff for Personnel, I know that many of the decisions we have made, and will make, are based directly on the results of your efforts. Your contributions have had a significant impact on the way we manage people; only the most naive would deny that we've been able to build a far better military establishment with you than we could have without you.

For these reasons, then, I welcome the opportunity to pass on some of my thoughts and, hopefully, to offer some challenges for the future. I'll start with a couple of observations.

In reading over past conference reports, I find that your guest speakers can be broadly categorized into two groups. In the first category are the personnel researchers, the experts, people like you. They understand the behavioral sciences. They understand the problems and the lead times involved in achieving effective progress. As a result, they usually tell an audience like this to ride out the changes in administrative personnel, to avoid a preoccupation with the operational demands of today's job, to take the long range view in the interest of excellence and real progress.

In the second category, I find people like myself, managers, or laymen (in the research vernacular) trying to cope with the day-to-day problems of procuring and managing large numbers of people. While we realize the futility of trying to come up with airtight solutions, we tend to ask you for formulas and tools (what you might call gimmicks) that will enable us to tie up our problems in nice, neat packages. And, of course, we're always in a hurry. We become a little impatient when your answers are not quick enough or you try to qualify, perhaps justifiably, the answers you do provide.

From experience, I know that the difference in perspective represented in these two points of view can put the manager and his experts at odds. I would go even further. They can, and often do, make effective communication difficult, if not impossible. The tragic thing is that we have never needed effective communication between the researcher and the manager more than we do today. The demands on our ingenuity to employ, to challenge, to get maximum value from those who elect to join us have never been greater. There is too much to know, too much to comprehend, too much that needs to get done, to waste time with jargon and petty bickering.

At this point, let me share a few personal reflections I believe to be crucial to understanding the nature and scope of the challenges that confront us.

I think that any assessment of our current situation has to begin with the basic premise that as a nation we are passing through some very tough times, perhaps the toughest in our history. It is a period characterized on one hand by questioning and uncertainty and on the other by an awakened social awareness in almost every sector of society. Institutions, as well as basic values and standards, are being challenged. Change and turbulence have become the rule rather than the exception.

You can attribute what you read and hear to a lot of things. But whether you blame the current climate on change, disenchantment with the "establishment," the war, the volunteer

effort, or some other factor, the fact remains that today's environment has some profound implications for the armed forces, especially when it comes to manpower.

We can talk about this situation in terms of tight money. We can point to force reductions, or shortfalls in military strength. What it all means, however, is that the military, as an institution, must face up to the challenge of having to manage its human assets better -- to get as much dedication, as much effectiveness, out of our people as possible. In spite of the present environment, we must continue to find legitimate ways to attract the manpower we need to keep our forces viable.

It's been estimated that maintaining an all-volunteer force of 2.2 million over the long haul will require one out of every three qualified and available men in this country to volunteer for active military service. Under present policies, one-third will have to be "above average," over one-half, "average." This year alone the armed services will have to attract better than 400,000 young Americans to meet our commitments. But the problem goes beyond numbers. They also have to be the right kind of people. Here is where we've got some room for improvement.

For example, out of the average 100 enlistees who enter the Air Force, 25 leave before they complete 4 years. Some are found unsuitable or unfit for military service; some have disciplinary problems; others leave for personal reasons. In addition to the personal anguish involved, these people represent time, effort, and money, all of which are in short supply these days. We know that high school graduates have fewer disciplinary problems than non-graduates, by a ratio of one to four. (Yet we know that the majority of non-graduates can do well.)

The question is why? What are the factors working in this area? Can we, can you, identify them? Is it something we're doing? Should we try to deal with these people before they come on board or after, or both? These are tough questions; but they need to be addressed, and I look to you for the answers.

After we get the right kind of people, we also have to manage them properly. I sincerely believe young people today are as eager as ever to do a good job. I also know they want to ask questions. They want to know why as well as what; and, above all, they want to be challenged and provided the opportunity to participate in decisions that affect their future.

This means we have to avoid arbitrarily poking these people into holes. We have to realize and act on the premise that what may excite and challenge one man or woman may not challenge another. In the long run, arbitrary, "convenient" actions can create a situation in which an individual who might have been a tremendous success ends up being bored and indifferent to his work or elects to leave us.

Our need here is obvious. We must have the best tools we can get to classify and assign people. We must have the best predictors of success in training and on the job that you can give us. We must be as sure as we are able that the system is doing everything it can to provide each person who comes to us the best opportunity possible to grow, and learn, and progress to the limits of his or her capabilities.

I would like to underline the fact that this philosophy applies to the minority as well as the majority members among us.

Like every other service, the Air Force is trying hard to satisfy the expectations of its minority members. We have almost 250 enlisted career fields in the Air Force. A little better than a year ago, 91 of them had less than a 5 percent minority population. Today that number is down to about 45, and we are heading for 0. Like the other services, we are working the problem and the effort is starting to pay dividends. We still can use help, all the help you can give us.

What I have been discussing so far is just a small fraction of the challenge we face. Even this incomplete list, however, indicates the need today for aggressive, decisive action. This is the time for the military to question, to search, to find new and better ways to manage its human resources, to take the lead in making its own unique contributions to the growth and development of military people as vital members of society.

We must ask ourselves, What are the implications for testing? What role should tests play in this coming decade? Of all the many specific uses to which tests are put, is there any one direction or unifying theme? I believe there is. I think there is a clear direction in which testing should develop, and I think there is an underlying theme for that direction.

It seems to me that the requirement to get maximum value out of our people -- all our people -- and the new emphasis on individual challenge and choice dictate that we move testing out of our impersonal "test 'em and tell 'em" mode of the last generation. The change, in my view, is long overdue. Whatever reasons required managing people almost solely by testing in wartime or periods of rapid mobilization, the competitive demands of a free-choice, peacetime labor market are different and they require different solutions.

Which means what, specifically?

First, I think it means that tests of the future must focus much more on the individual, on his unique talents and desires.

Second, it means fitting tests into a scheme that maximizes the opportunity for choice, both by the individual and by his employer.

Third, as a function of the first two, it means that tests must provide broader profiles of information. Four or five aptitude scores do not provide a sufficient basis for the kind of complex, comprehensive career programs we have today and will need in the future.

The role of tests, now and in the future, should be to supply a vital piece of the information we need in properly managing the careers of our people. The quality and breadth of this information will be crucial to the quality of the decisions we make, but tests must not serve to prejudge the individual for a full career. They should clearly be a tool in the decision process. By that I mean that they should show the implications of various alternative decisions.

For example, tests are needed, especially with the advent of the new instruction systems, that can recommend specific kinds of learning experiences that can be applied to an individual case. Lee Brokaw, your program chairman, put it well recently when he said, "We must shift emphasis from finding the best individual for the job to find the best job for the individual."

New systems are also needed to provide a more balanced role for tests, roles that clarify where tests fit in career development, roles that make sure the limitations of testing are equally clear.

We also have to remember that testing policies and practices must have a relationship to this nation's goals in the area of social policy. The Civil Rights Act of 1966, the Supreme Court decision in the Duke Power Company case, and the Equal Employment Opportunity Act of 1972 all really make one point. Equal opportunity means giving all Americans a chance, and this chance must be more than just a chance to fail.

As Lyndon Johnson said in the last address he made before his death, "History has not been equal for all Americans." We recognize now that tests can be used negatively. They can perpetuate a cycle of assignment to jobs with limited growth potential. They can confirm the "inequality of history," in Lyndon Johnson's terms, by making past difficiencies the basis for continued unfairness. In short, they can become sentences instead of doors to wider growth and development.

Ladies and gentlemen, if we are to have viable armed forces in the future, we must come up with reasonable answers to the questions and problems confronting us in personnel management. You know our requirements; you know how urgent they are.

Today, as your keynote speaker, I ask you to provide us the tools we need to satisfy these requirements. When you bring us the tools, tell us how to use them and how not to use them, and make sure we listen. As a manager with a primary interest in people, I can tell you we do need help, the kind of help only you can provide. I ask that you come forward and tell us what should be done.

For the short term, give us the best answers you can. Remember, we can't always wait for the research effort to be neatly organized, the right researcher to be found, the research effort to be conducted in a simon pure environment, etc. Very often we need answers now. I challenge you to respond quickly when the situation demands.

For the long term, identify the areas where we should encourage further research: research that is relevant; research designed to tackle the tough problems of tomorrow, not yesterday's problems; research that will enable us to achieve progress that is real and enduring.

And finally, let the experts and the managers (you and I) join forces, with the knowledge that there is immediacy and urgency in the work we do, that there is excitement and opportunity in improving the management of the most precious asset entrusted to any organization. Now may not be the best of times, but it is our time. Let's use it wisely.

Synopsis

MTA Symposium - Cooperative Aptitude Research Programs Conducted through the Armed Forces Vocational Testing Group

Participants:

Dr. Harry D. Wilfong, AFVTG
Maj Verna S. Kellogg II, AFVTG
Capt Melvin T. Gambrell, Jr., AFVTG
Dr. Lonnie D. Valentine, AFHRL - Chairman

Introductory Comments (Dr. Valentine)

1. The Armed Services High School Testing Program has gradually evolved since 1966 into a program of sizable importance.

2. In today's all volunteer era, it will probably assume greater importance to the services as a constructive point of contact between the services and the educational community with its pool of young people entering military age.

3. For about the past 6 months, the High School Testing Program has been under management of the Armed Forces Vocational Testing Group at Randolph Air Force Base. Other participants in today's symposium are from that organization.

4. After the participants have each briefly discussed some aspect of the program, the floor will be open for audience questions and suggestions. It is hoped that the discussion period will serve to elicit constructive suggestions for the program.

Program History (Major Kellogg)

1. A popular stereotype of the recruiter is the "lifer" body-snatching young men from the ghetto and the juvenile courts and channeling them to the rehabilitation environment of the military services. Nothing could be further from the truth. The services need "quality" human resources--military services today demand trainable, educable and alert individuals.

2. The sophistication of today's weapon systems requires the military services to expend more time and money training a man to an acceptable skill level than they can anticipate realizing from his services. Thousands of young people leave the military service after their first enlistment. Return on the taxpayers' investment in

them is realized mainly in terms of the skills they take back into civilian jobs.

3. Because of this requirement for quality manning, the military services have tied their recruiting and enlistment programs directly to the measurement of aptitude for vocational training.

4. The Department of Defense, this year, directed the formation of the Armed Forces Vocational Testing Group--to direct, manage and guide the Department of Defense High School Testing Program. This program is the cornerstone of recruiting the all-volunteer force.

5. The primary tool of the testing program is the Armed Services Vocational Aptitude Battery, commonly referred to as ASVAB.

6. ASVAB is the product of over 30 years research and development in the area of military classification--most of it accomplished by people in this room. Today's enlistee is carefully tested for identification of his potential for training.

7. The genesis of the High School Testing Program goes back to 1958, when the Air Force recruiters struck on the idea of going into high schools with a Vocational Aptitude Battery, enabling them to identify specific youngsters with specific aptitudes as potential recruits. With the success of the Air Force Program, the other services initiated similar programs. The competition caused confusion and projected an unprofessional image of the military services.

8. In 1966, the Department of Defense directed that the services combine their high school efforts.

9. A group of service behavioral scientists met and the eventual fruit of their efforts was ASVAB--ASVAB combined the best and most desired features of all the military classification tests and gave the High School Testing Program a product usable by all the services.

10. The concept was good--but different parts of the action were assigned to the various services--the Army was tasked to develop the test, the Navy was tasked to provide counselor's materials and the Air Force supervised the scoring and return of the test to the counselor. Recruiters went back into the schools, all using the same test, but still with their parochial interests at heart. Thus, ASVAB became an Army test, a Navy test or an Air Force test depending upon the representative who got there first.

Both recruiters and counselors lost sight of the value of the battery in helping young people plan their futures--and got hung up on the idea that ASVAB was only a means to get names of potential enlistees.

11. The Armed Forces Vocational Testing Group is the answer to the problem. Staffed by personnel of all the military services--Army, Navy, Marines and Air Force--the group provides a single manager and interpreter of ASVAB for the Department of Defense High School Testing Program. Dealing with the local recruiter through an interservice recruitment committee (organized by geographical location and representing each service) the group assures equitable and professional administration of ASVAB.

12. The mission of the Armed Forces Vocational Testing Group is to place ASVAB in the schools, assure its currency and validity and provide the student and counselor with assistance and information. Our job ends there; if the youngster is interested in the military as a means to vocational training, the recruiter provides him with that information after the test has been administered.

13. Overt recruiting is not allowed in conjunction with either the administration of the battery or the delivery of the scores to the counselor and the student. Improved recruiting is an anticipated side effect in those instances where aptitude and service interest are demonstrated.

Armed Services Vocational Aptitude Test (Capt Gambrell)

1. Over the past 30 years, the various military services have used aptitude tests to select, classify and place enlistees. As Major Kellogg has previously mentioned, in 1958, the Air Force conceived the idea of going directly to the high schools with an aptitude test, the AQE. This proved to be a most successful program for pre-qualifying prospects for enlistment. The other services soon adopted similar programs and high school aptitude testing by military services became a reality.

2. I would like to direct my section of this panel to a review of development of the ASVAB, content and standardization. The primary purpose of this review is to reacquaint you with the basic structure of our current ASVAB.

3. The initial battery development study (Bayroff & Fuchs, 1970) was directed toward (A) identifying interchangeable tests within the existing service classification batteries, (B) developing

tests for the joint battery from a pool of test items composed of those in the interchangeable components of the existing service batteries and (C) standardization of the resultant battery.

4. To establish test interchangeability, all of the service classification batteries were administered to a common sample of 3900 basic trainees from the Army, Navy, Air Force and Marine Corps. Testing was arranged for three separate days, with only one of the service batteries being administered on a given day, and battery administration was counterbalanced to eliminate effects of test practice.

5. A stratified subsample, representative of the full range of ability on the Armed Forces Qualification Test, was used to establish intercorrelations among all of the tests in the various service classification batteries; this resulted in the identification of seven interchangeable content areas. Two other areas were added to ASVAB to yield an AFQT equivalent.

6. The nine ASVAB tests yield five aptitude composites which are reported to high school counselors through standard format. These are:

(A) Electronics - describes students in terms of abilities relevant to electrical/electronic occupations. The aptitude score consists of tests dealing with electrical information and understanding mechanical principles.

(B) General Maintenance - describes capabilities relevant to a variety of mechanical and trade jobs. The composite consists of tests assessing shop information and spacial ability.

(C) Motor Maintenance - is concerned with ability for engine repair and related jobs and measures automotive information and understanding of mechanical principles.

(D) Clerical - measures ability relative to clerical/ administrative occupations. Composite concerned with verbal ability and clerical speed and accuracy.

(E) General Technical - describes students' ability for occupations requiring academic ability. This aptitude is taken from verbal and mathematical components of battery.

7. The battery contains 300 test items; 100 in the Coding Speed and 25 in each of the remaining eight tests. Actual testing time is 112 minutes; generally two and one-half hours are required to administer the battery. Speed is not emphasized, although all

tests are timed.

8. Test items contained in Form 1 of the ASVAB (with the exception of those in Coding Speed) were selected from an item pool consisting of all items contained in the service classification batteries used in the "interchangeability" study. Criteria for item selection were difficulty level (proportion of examinees responding correctly), discrimination level (ability of the item to discriminate correctly between individuals who score high and those who score low on the relevant ability as reflected in the item's correlation with other items of its type), and content validity.

9. The 25 items in each subtest (except coding speed) were arranged in approximate difficulty order so that about 85 percent of examinees would answer the first few items correctly, with the success percentage decreasing to about 25 percent at the end of each subtest.

10. The first form of the ASVAB was standardized on stratified sample tested at a geographically representative sample of 11 Armed Forces Examining and Entrance Stations (AFEESs). The stratified sample was designed to be representative of the scores obtained by the mobilization population on the AFQT.

11. Forms 2 and 3 of the ASVAB were developed from a pool of new items, but procedures followed in their development and standardization are very similar to those for Form 1.

12. The reliabilities and validities of the battery have been found to be quite satisfactory. Kuder-Richardson reliabilities range from .78 to .87. The validity coefficients for service technical school courses are about equal to those of the earlier classification batteries.

ASVAB Research (Dr. Wilfong)

1. Since the history, development and content of the ASVAB have already been thoroughly covered by Major Kellogg and Captain Gambrell, I shall limit my comments to a brief description of research, both on-going and projected, conducted in support of the program.

2. Our immediate efforts are directed toward increasing the usability and application of the ASVAB counselor's manual. Toward this end--the normative base is being expanded and a broader sampling of military/civilian occupations will be included in Volume 2.

3. From the historical tape on ASVAB maintained at AFHRL, Lackland AFB, the Personnel Research Laboratory has extracted a

representative sampling of 36 thousand cases producing separate norms by geographical region, sex and high school grade. It has been my experience, thus far, that these are the three most common comparative variables requested by high school counselors.

4. The other ASVAB research requirement is to expand the listing of civilian/military occupations as cataloged in Volume 2 of the counselor's manual. What is being planned - assuming sufficient lead-time to permit publication for the 1974-75 school year--as a two-columinar hierarchical listing of representative civilian and service jobs anchored to appropriate dictionary of occupational titles numbers.

5. Long-range research needs identified in support of the ASVAB program are too numerous to cite during this brief presentation. However, I would like to touch upon some of the more significant efforts that are planned.

6. First, content revisions are being considered, including deletion of the tool knowledge subtest, possible shortening of some of the other portions of ASVAB and the insertion of an occupational interest schedule. These changes should appear in Forms 5 and 6 of the ASVAB, hopefully, to be available by the 1975-76 school year.

7. It is the general consensus of researchers exposed to the ASVAB that a highly desirable research undertaking would be development of current mobilization base data--representative of modern America on as many strata as possible. What is envisioned is a study broad enough to serve many masters and the needs of many agencies while simultaneously substituting for the old 1942-44 mobilization data base. The many problems involved in planning, securing approval for, and executing such an ambitious undertaking are well recognized.

8. Another, equally critical, avenue for research exploration is to establish a mechanism for conduct of a series of longitudinal validation studies tracking students at successive stages along their career path, both within and outside of the military service. By its very nature, the ASVAB program presents the ideal vehicle for performing such research.

9. A wide scenario of research support requirements has been identified for the ASVAB program, including subjects as listed in Attachment 1, Armed Services Vocational Aptitude Testing Research and Development Support requirements, of the joint services regulation covering the Armed Forces Vocational Testing Program.

10. Research needs identified by high school counseling personnel and by service recruiting, production and classification representatives range across a broad spectrum, covering such topics as: tracing the longitudinal career decision process; exploring the feasibility of remote test scoring and concurrent feedback; establishing a task/aptitude criterion matrix, permitting, in turn, rapid estimation of predictive validities for new jobs entering the inventory; reconfigured jobs, or jobs having an insufficient N for traditional validation purposes; and a linear programmed approach to test administration.

11. Rather than describe the entire spectrum of our anticipated research objectives at this time, I would prefer to return the floor to the workshop chariman so that we can explore subjects in which you may have a particular interest.

Concluding Comments (Dr. Valentine)

1. My few comments represent my personal views--not necessarily policy of AFHRL or the Air Force.

2. AFHRL's role, as well as that of the other service R&D labs, in the program is viewed as one requiring sound R&D to support and improve the program. Policy with regard to the program is the responsibility of service and DOD Hq agencies.

3. There are three areas in which substantative support R&D is needed. These deal with (a) breadth of battery content and coverage, (b) battery standards (which include both norming and composite structure), and (c) validity studies.

4. In its present form, ASVAB is limited to subtests either "common" to earlier service classification batteries or necessary to provide an AFQT score. Most of these predecessor batteries were designed to provide optimal prediction of training success. Such composites tend to intercorrelate substantially. School counselors have always been constrained to counsel all students, and the optimal test system for them both identifies ability level and provides optimally different classification indexes. The services have been able to select rather than to simply classify in the true sense. Under the volunteer force the services are going to be compelled to join counselors in a more effective job of differential classification. In addition, the ASVAB's present limited content coverage short-changes some of the classification needs of some of the services. Thus, the service labs need to give careful attention to appropriate added content and to the classification models which will allow decisions about both level of aptitude and the differential aptitude demands of jobs.

5. The mobilization population standards against which the services typically calibrate new tests is now 30 years old, and it does not take into account special subgroups of particular interest for special programs. A broad based national normative census, with provision for identification of standards appropriate to particular population subgroups, is needed. Some counselors have expressed dissatisfaction with the service's 1942 mobilization base standard and have maintained that they want standards based on high school age youth.

6. Substantative ASVAB validation studies not only against service criteria but against civilian career criteria as well, are needed.

7. Careful attention to these three areas, with cooperative planning and research by the various service labs, can lead to development of a comprehensive and flexible test battery capable of satisfying both service and counselor needs, and can be an invaluable recruiting aid to all the services. It's accomplishment will, however, require that the services lay aside parochial interests, support the appropriate R&D labs with the manpower and other resources needed for its accomplishment, and encourage long range cooperative planning and research.

(The symposium was concluded with an audience question and reaction period.)

DEVELOPMENT OF THE ARMED SERVICES VOCATIONAL APTITUDE BATTERY

M. A. Fischl, Ph.D.
U. S. Army Research Institute for the Behavioral and Social Sciences

The Armed Services Vocational Aptitude Battery (ASVAB) is the test battery administered in the joint services high school testing program. In this program the services offer to administer a battery of vocational aptitude tests in any high school that requests it. Trained test administrators visit the school and administer the battery, the tests are scored centrally, and results are returned to the high school guidance counselor for transmittal to the student. This program is provided to high schools that request it, at no charge either to the school or to the student.

Identification of Content

The tests employed in this program were developed to be a set of short alternate forms of the common elements of classification test batteries in use by the services. At the time of the original development each of the services was administering a classification battery of nine or ten aptitude tests. To identify the common content material of these batteries, the batteries of the Army, Navy, and Air Force were administered to a national sample of 3900 enlisted men during their first few days of service. Each enlisted man took the tests of all three batteries, in a counterbalanced sequence, one battery a day, over a several day period. The sample was subjected to statistical adjustment to equate for differences in screening methods and acceptance standards among the services, and the test scores were intercorrelated to determine which test content areas were sufficiently similar as to comprise the core of a joint test battery. The criterion of similarity

was set as a correlation coefficient of 0.90[1/], and this analysis identified the seven test content areas of verbal fluency, quantitative reasoning ability, spatial visualization, mechanical comprehension, shop information, automotive information, and electronics information. An eighth content area, clerical ability, was added because of its importance for counseling and classification, even though the service tests of this function did not meet the statistical criterion. Finally, knowledge about hand tools was added as a content area so that all four areas assessed in the Armed Forces Qualification Test (AFQT) would be represented in ASVAB.

Construction of the Original Tests

For all but the clerical content area, 25 items, spanning a full range of difficulty, were selected from the operational tests of the services. These items were assembled so as to comprise new forms of the service tests in the respective content areas. The clerical ability test was assembled by taking, intact, the most valid of the service tests in this area. It is a 100-item speed test; all of the others are power tests and are, with the exception of Tool Knowledge, half-length forms of the service tests in the counterpart areas. The Tool Knowledge test contains the same number of items as was contained in the Tool Knowledge portion of the then operational AFQT.

Exhibit 1 describes the contents of the nine tests of the battery. To obtain norms, these tests were administered to a national sample

Insert Exhibit 1 about here

1/ Corrected for restriction of range and lack of perfect reliability.

EXHIBIT I

TESTS IN THE
ARMED SERVICES VOCATIONAL APTITUDE BATTERY

1. CODING SPEED TEST (CS). IN THIS TEST THERE IS A KEY AND 100 ITEMS. THE KEY IS A GROUP OF WORDS WITH A CODE NUMBER FOR EACH WORD. EACH ITEM PRESENTS ONE WORD FOR WHICH THE EXAMINEE INDICATES THE CODE NUMBER.

2. WORD KNOWLEDGE (WK). EACH ITEM REQUIRES THE EXAMINEE TO SELECT THE CORRECT SYNONYM FOR A SPECIFIED WORD.

3. ARITHMETIC REASONING (AR). EACH ITEM IS A REASONING PROBLEM INVOLVING APPLICATION OF ARITHMETIC PROCESSES.

4. TOOL KNOWLEDGE (TK). EACH ITEM PRESENTS FIVE DRAWINGS OF VARIOUS TOOLS OR SHOP EQUIPMENT. THE EXAMINEE INDICATES WHICH OF THE FOUR ALTERNATIVE DRAWINGS GOES BEST WITH THE LEAD DRAWING.

5. SPACE PERCEPTION (SP). EACH ITEM CONSISTS OF 5 DRAWINGS: A PATTERN AND FOUR BOXES. THE QUESTION TO BE ANSWERED IS WHICH ONE OF BOXES CAN BE MADE BY FOLDING THE GIVEN PATTERN.

6. AUTOMOTIVE INFORMATION (AI). EACH ITEM ASKS A QUESTION ABOUT THE IDENTIFICATION OR OPERATION OF AUTOMOBILE PARTS.

7. SHOP INFORMATION (SI). THIS TEST HAS QUESTIONS ABOUT SHOP PRACTICES AND THE USE OF TOOLS. MANY OF THE ITEMS CONTAIN DRAWINGS.

8. MECHANICAL COMPREHENSION (MC). EACH ITEM INCLUDES A DRAWING OR DRAWINGS ILLUSTRATING SOME PHYSICAL PRINCIPLE AND A QUESTION CONCERNING IT.

9. ELECTRONIC INFORMATION (ELI). THIS TEST HAS QUESTIONS ABOUT ELEMENTARY PRINCIPLES OF ELECTRICITY AND ABOUT ELECTRICAL OR ELECTRONIC DEVICES.

of 3050 Selective Service registrants, stratified to represent the population of young men of military age.

Construction of Successor Forms

Coincident with introduction of the original form of the battery in the school year of 1968-69, work began to develop two entirely new forms. For each of the eight power tests[1/] six times the required number of items were prepared. This is customary procedure, preparing a substantial overage in anticipation of attrition in the item analysis stage. These items were administered to several national samples stratified to represent the population of military age young men. The total n was 4000 cases, 18% of whom were blacks, and item analysis statistics of difficulty and homogeneity were obtained. Using these statistics, items were assembled into two parallel 25-item forms of each of the eight power tests. Each test was constructed to be of wide ranging difficulty, with items spanning virtually the entire range of 99 per cent difficulty (very easy) to 1 per cent difficulty (very hard) in the reference population. When coupled with the new forms of the speed test, there were two entirely new nine-test batteries.

These batteries were administered to several stratified national samples of Selective Service registrants, one form to a sample. The total n was 3500 cases, virtually all whom were between 18 and 19 years of age, 80% of whom had completed between 10 and 13 years of education, and 15.5% of whom were black. From this administration percentile and Army Standard Score norms were developed, and test reliability coefficients, intercorrelations, and other characteristics were derived.

1/The speed test was prepared in its final forms, because item analysis of a speed test is not meaningful, and item attrition is not incurred.

Characteristics of the Battery

The entire battery requires approximately 2 1/2 hours to administer. The longest test, Arithmetic Reasoning, takes about 25 minutes; the shortest, Coding Speed (clerical aptitude), takes 7 minutes; most of the others are 10-minute tests.

Reliability coefficients are presented in Table 1, and test intercorrelations appear in Table 2. The intercorrelations are of intermediate

Table 1

ASVAB Reliability: Coefficients of Equivalence and Internal Consistency for Each Test

Test	Equivalence 1/	Internal Consistency 2/
Coding Speed	.83, .86	Not Applicable
Word Knowledge	.80, .85	.87
Arithmetic Reasoning	.81, .86	.87
Tool Knowledge	.79, .76	.78
Space Perception	.82, .84	.84
Mechanical Comprehension	.77, .73	.83
Shop Information	.67, .65	.81
Automotive Information	.75, .74	.85
Electronics Information	.80, .79	.83

1. Correlation of Form 2 with Form 3, in two subsamples, Form 2 administered first in one (n = 536), Form 3 administered first in the other (n = 514).

2. Utilizing Kuder-Richardson Formula 20, n = 616.

(insert Table 2 here)

TABLE 2

INTERCORRELATIONS OF ASVAB TESTS, N = 610

TEST	CS	WK	AR	TK	SP	MC	SI	AI	EI
Coding Speed (CS)									
Word Knowledge (WK)	.58								
Arithmetic Reasoning (AR)	.56	.72							
Tool Knowledge (TK)	.27	.38	.36						
Space Perception (SP)	.39	.58	.62	.50					
Mechanical Comprehension (MC)	.45	.67	.68	.58	.69				
Shop Information (SI)	.35	.50	.50	.69	.56	.71			
Automotive Information (AI)	.35	.47	.49	.62	.46	.67	.76		
Electronics Information (EI)	.44	.64	.61	.58	.61	.73	.69	.69	

magniture, with the perceptualmotor test (Coding Speed) being generally most independent of the rest of the battery.

The reliability coefficients seem reasonably high for tests of only 25-item length, and predictive accuracy is enhanced operationally by utilizing combinations of several tests. Why the Shop Information test yielded the lowest equivalence coefficients in the battery is not apparent, especially since its internal consistency coefficient is of comparable magnitude to the rest of the set.

Occupational predictions from ASVAB are generally made utilizing combinations of several tests. These combinations, both with ASVAB tests and their counterparts in the services' operational batteries, have undergone extensive validation study against criteria of training success in large numbers of job specialties. The modal validity coefficient against these criteria has been of the order 0.60, with a range of 0.35 to 0.75.

In summary, the ASVAB is a nine-test aptitude battery developed to be shortened alternate forms of the common elements of the then service classification batteries. It has been empirically developed, and tried out on large representative samples of the population to which it is intended to be applied. The original form was in use in the joint service high school program for five years, Form 2 has been operational in the schools since the beginning of this school year, and testing goals are in excess of a million students by the end of the year.

"NEW DEVELOPMENTS IN DEFENSE LANGUAGE PROFICIENCY TESTING"

PRESENTED BY DR. C.D. LEATHERMAN AND DR. A. AL-HAIK OF THE
DEFENSE LANGUAGE INSTITUTE

1973 MILITARY TESTING ASSOCIATION CONFERENCE, SAN ANTONIO, TEXAS
28 OCT - 2 NOV 1973

PURPOSE

Defense Language Proficiency Tests, commonly referred to as DLPT's are tests designed to measure the language proficiency of all DLI graduates, as well as DOD-connected individuals claiming proficiency in a particular foreign language. Specifically, the DLPT serves two major purposes:

a. To measure the listening comprehension and reading comprehension skills in a foreign language. These two skills range from minimum ability which may be useful in an activity such as directing traffic, to the skill needed by a competent interpretor. The test does not discriminate among the higher language skills required by a top-level simultaneous translator.

b. To evaluate and rate the examinee's ability to meet the linguistic requirements for particular jobs, such as interpretor, interrogator, and translator. The test is not intended to assess other skills, such as interrogation techniques, which may be required by these jobs.

BACKGROUND

The history of military proficiency testing goes back to 1948 when the Army first developed what was then called The Army Language Proficiency Tests. Tests in 31 languages were developed and used between 1948 and 1953. Each test consisted of four parts: Parts I and II covered understanding of the spoken language; Part III dealt with reading ability; and Part IV measured writing ability through knowledge of grammar. Due to military pressures at

the time, the development of these tests involved only a minimum of research. During and following the Korean War it became apparent that the tests were not always reliable in discriminating between individuals of varying levels of language ability. As a result of this recognized deficiency, Army user agencies requested the development of new Army Language Proficiency Tests. Accordingly, in 1954 The Adjutant General's Office was directed to develop two prototype Army Language Proficiency Tests: one for an Asiatic language, namely, Chinese-Mandarin and one for a European language, Russian. Using these tests as guides the Army Language School (the predecessor of The Defense Language Institute) was subsequently directed to prepare language tests, both to replace existing instruments and to develop new tests in languages for which there were no tests at that time. Even though the new prototype tests in Russian and Chinese-Mandarin were carefully developed and validated, the other new tests, approximately 38 in number, consisted mainly of translations of the two prototype tests and were not subjected to systematic validation procedures before their introduction into the Defense Language Program in 1958. As a result, Russian proficiency test was used as the model for all European languages and the Chinese-Mandarin test was used for all Asiatic languages. It became apparent following extensive use during the subsequent years that these 38 additional tests were not as valid or as reliable as had been assumed. Additionally, the use of a single test form for each of these tests over more than a decade further reduced their effectiveness due to possible test compromise. Recognizing these deficiencies, The Defense Language Institute initiated a series of new projects in 1966 aimed at developing follow-on DLPT IIs, or second generation DLPTs to replace the old ones. The first group of DLPT IIs, were developed under

contract with the Educational Testing Service, utilizing language experts from the DLI and from various universities, in addition to resources from ETS. Seven high-enrollment languages were included, such as Russian, Spanish, Arabic, etc. These initial DLPT IIs are introduced into the Defense Language Program at the present time. As these tests are introduced into the system and prove to be operationally sound, other language proficiency tests and alternate forms will then be introduced into the system. Current plans call for the development of two or three alternate forms as a minimum for each high density DLPT II. DLI's R&D plan includes the construction and validation of a computer bank of test items for each language, from which multiple test forms can be assembled automatically and actually produced in final form, both test booklets and audio tapes, by computer. These were the facts and some of the events which led to our current efforts with regard to second generation DLPTs. Additional details concerning specific aspects of the developmental work program are given below.

DEVELOPMENT OF PROTOTYPE TESTS IN RUSSIAN AND CHINESE-MANDARIN

These early prototype tests were developed by Army Language School personnel, in cooperation with representatives from the Personnel Research Branch of The Adjutant General's Office. Each test has two parts: Listening Comprehension, consisting of sixty items and Reading Comprehension also consisting of sixty items. All items are multiple-choice type, each consisting of a lead sentence or a short paragraph in the foreign language and four optional responses printed in English in the test booklet. Three of these options are plausible distractors whereas, only one is the correct answer. The options can be words, phrases, sentence translations, or paragraph gistings depending on the objective of the item. Listening comprehension item leads are voiced on a tape,

but reading comprehension item leads are printed in the test booklet. After completion, the tests were field tested, including a validation measure, with Army personnel possessing varying degrees of skill and experience in the two languages. The validation instrument consisted of job-oriented tasks such as interpreting and translating. The median correlation between the Chinese test and the criterion measure was .70 and between the Russian test and the criterion measure .78. The K-R 20 reliability of the Chinese test was .96 and the Russian test .95.

DEVELOPMENT OF THE FIRST GENERATION OF PROFICIENCY TESTS IN 38 LANGUAGES

Language specialists at The Army Language School were furnished English translations of the two prototype tests mentioned previously and were then instructed to use these translations, item by item, in developing the new tests. Since the conditions for deviating from the given translations were few and rather rigid, the resultant new tests were essentially parallel in many ways to the prototypes. Unfortunately, not only were the options identical, but the correct option remained as the same key across all tests. It was subsequently discovered that this procedure increased the chances for test compromise, much more than usual. The reason is that many DLI graduates would either subsequently reenroll in a different language course or take a DLPT to establish military competency in a foreign language studied elsewhere. Furthermore, an erroneous assumption was more than likely made at the time that since all the tests were parallel in content and format, they would therefore all have the same level of difficulty. For this reason, no attempt was made to standardize raw scores and the same skill-level cut-off scores were used across all languages. This initial assumption was unfortunate and was incorrect as later validation data disproved the theory.

DEVELOPMENT OF SECOND GENERATION DEFENSE LANGUAGE PROFICIENCY TESTS (DLPT II)

It was considered essential to correct the operational situations described above and to replace the then current tests which were outdated and possibly compromised through the overuse of only one test form per language. Accordingly, second generation DLPTs are being developed, printed and systematically introduced into The Defense Language Program during a three to four year period. The new second generation tests have the following characteristics:

a. Although they are identical in format to the old test, the content of each is unique and independent from any other. Test content is based strictly on a description of skill-level objectives for each language.

b. Current plans specify at least three alternate forms for high-density languages and two forms for low-enrollment languages.

c. All test forms will be subjected to item analysis and form-equating procedures.

d. The validation of each test will be based on the collection and analysis of field data concerning job-performance ratings.

e. Each newly developed test will be independently normed and new skill-level cut-off scores will be established.

f. Raw scores will be converted into standard scores to achieve mutual comparability for languages which have varying degrees of difficulty.

PROJECTED PRODUCTION SCHEDULE FOR SECOND GENERATION DLPTs

DLI R&D PRIORITY I (DLISDA, P-0015)

(CM)	Chinese (M)	(LA)	Spanish (LA)
	Arabic (Modern Standard)*	(TH)	Thai
(AE)	Egyptian	(VN)	Vietnamese (N)
(AP)	Syrian	(VN)	Vietnamese (S)
	Iraqi	(RU)	Russian
	Saudi		

These DLPT IIs, Audio Tapes, DLI PAM 611-1, OPSCAN answer sheets are scheduled for distribution world-wide during 1st and 2nd Qtr, FY 1974.
*Two Arabic audio tapes (AE, AP) will be used with the Arabic test booklet.

DLI R&D PRIORITY I (DLISDA, P-0015-2)

(FR)	French	(KP)	Korean
(GM)	German	(CM)	Chinese-Mandarin (Simplified characters)
(JA)	Japanese	(YD)	Chinese-Amoy (originally Priority III)

Two of these DLPT IIs have been completed. The remaining three are scheduled for completion during 3rd Qtr, FY 1974. Publication and world-wide distribution is scheduled during FY 1974.

DLI R&D PRIORITY II (DLISDA, P-0016)

(BY)	Burmese	(PL)	Polish
(CC)	Chinese-Cantonese	(PQ)	Portuguese Brazilian
(CX)	Czech	(TU)	Turkish
(HU)	Hungarian	(CA)	Cambodian
(IN)	Indonesian	(LC)	Laotian

These DLPT IIs are programed for completion during 3rd and 4th Qtr, FY 1974.

DLI R&D PROTOTYPE TEST DEVELOPMENT (PRIORITY III AND IV) (DLISDA P-0055)

(HJ)	Hindi	(HS)	Hausa
(AC)	Amharic	(HC)	Haitian-Creole

These are prototype tests designed to resolve a deficiency in that no DLPTs are currently available. These test emphasize the evaluation of reading proficiency in the target language. They are to be used in conjunction with a "linguistic data sheet" in order to assess linguistic background factors not measured by the tests. Scheduled completion dates are 3rd and 4th Qtr, FY 1974.

DLI R&D PRIORITY III (DLISDA, P-0130-1

(AB)	Albanian	(PT)	Protuguese (European)
(BU)	Bulgarian	(RQ)	Romanian
(CM-M)	Chinese-Mainland	(SC)	Serbo-Croatian
(GR)	Greek	(SR)	Spanish (Castilian)
(JT)	Italian	(SW)	Swahili
(ML)	Malay	(UK)*	Ukranian (deleted)
(PF)	Persian (Farsi)		

Initial R&D work is underway; completion is expected during FY 1975.

DLI R&D PRIORITY IV

(AA)	Africaans	(NR)	Norwegian
(DA)	Danish	(PG)	Persian (Afghan) 1/
(DU)	Dutch	(SL)	Slovenian
(FJ)	Finnish	(SY)	Swedish
(HE)	Hebrew (Mod)	(UR)	Urdu 1/

Preliminary R&D work is underway; due to the lower priority a scheduled completion date is not feasible at this time. 1/ No DLPTs are available and structured interviews may be the more expedient alternative.

Milton H. Maier
paper presented
Military Testing Association
San Antonio, Texas
29 Oct - 2 Nov 1973

Enlisted Selection and Classification Testing in the US Army

Introduction

For several years now testing has been in a state of turmoil. Previously, tests were quietly given in schools, industry, and the armed services, and except for those who were directly involved, such as researchers and personnel managers, no one really paid too much attention to what was happening in the field. In the 1960's, though, testing, along with so many social practices and customs, was called into question, and today testing practices are regulated by law, the courts, and the Equal Employment Opportunities Commission, as well as by the test developers and users. Whereas in former times, a new test might be of local interest to the immediate users, or perhaps to the professional community if it had some unique characteristics, today a new test is subject to scrutiny by a good many parties. In the midst of this concern, the new Army Classification Battery, or ACB, was born, and as might be expected, its arrival did not go unnoticed. In my paper, I will describe briefly what the new ACB is, how it is used operationally, and finally give some comments on how it has fared since its introduction on 1 May of this year.

What is the new ACB?

The new ACB arrived on the scene at about the same time as the all volunteer Army, and it is admirably suited to the current needs for selecting and classifying enlisted personnel. One advantage is that the

new battery has a broader coverage than the old one. The two primary changes are that we expanded the number of subtests in the academic or general ability domain, and that we included more tests of interests and attitudes or measures of the willingness factor. We retained the old standbys of Arithmetic Reasoning and Word Knowledge, and we added measures of Science Knowledge and Mathematics Knowledge to provide more precise measurement in this critical area of general mental ability. By including these additional tests, we obtained more accurate prediction of success in the different job areas. (Chart 1)

The other major improvement is that we added more interest measures. The earlier ACB contained the Classification Inventory, which was a self-description instrument to help identify the infantryman. The original inventory grew out of research begun during the Korean War. We did not, however, have any other self-description measures in the ACB, until the new ACB was developed. Now, we still have a measure of combat interest, updated during the Viet Nam conflict, and three other interest measures in electronics and mechanical maintenance, and an interest in attending to detail, used for clerical-administrative jobs, field artillery, and for equipment operators, such as drivers. The interest measures add to the prediction of success in relevant job training, in addition to the measures of general ability and specific measures of information in a field, such as electronics information and automotive information.

The information tests also tap a willingness factor. They cover material that most anyone with an interest and aptitude in the area can pick up. Electronics and automobiles are common to our culture, and a young person can readily become expert in these areas if he has the initia-

tive to engage in relevant activities. And as part of our rapid social change, young women no longer are as excluded from these areas as in former times.

The Trade Information subtest covers content peculiar to the skilled construction trades.

Finally, we have two perceptual tests that do not involve any reading. These are the familiar pattern analysis or spatial ability, and a new test, called Attention to Detail, which requires the ability to discriminate the letter C embedded in a series of O's.

The tests are combined to provide nine aptitude area scores, which are used as prerequisites for job training. The nine areas cover the full range of Army jobs, from combat and field artillery, through the maintenance areas - electronics, general, and mechanical, to clerical administrative and a set of jobs called skilled technical, our medics, MP's, military intelligence, and the like. We have two other job areas that are more unique. One is called surveillance and communications, which includes radio operators and target acquisition jobs. The other is called operators/food, which includes missile crewman, cooks, and drivers. (Chart 2)

I should say that all of these job groupings are empirically based. We did not force the jobs to come out in any special way. We examined the profiles of validity coefficients both visually and analytically to determine which jobs clustered together, and the grouping shown here is what emerged.

Looking at the subtests in each aptitude area, it becomes apparent that Arithmetic Reasoning, AR, is omnipresent; it occurs in seven of the nine areas, plus in GT, which is used to determine eligibility for additional testing or special programs. Another fact is that the new aptitude

areas are more complex than the old ones. These contain three to five subtests each, as contrasted to only two in the old aptitude areas.

The result of putting so many tests in each composite is that the new aptitude areas are more accurate predictors of success in job training, which is good, but also that the scores are more highly intercorrelated, which at first seems bad. We found, however, that the new area scores do a much better job of sorting out the manpower into appropriate job areas than did the old ones.

We analyzed the new scores by subjecting them to extensive analyses through simulating scores on a computer to represent random samples from the mobilization population. We assigned the simulated "men" to job areas based on both the new and old aptitude areas scores. We kept the quota restraints that certain percentages had to go into each job area which reflected the real Army quotas at the time. The consistent result is that the performance estimates were higher under the new set of scores than under the old. The increase in predictive accuracy more than compensated for the increased intercorrelation.

Another gain realized from the new aptitude area is that each of them contain tests that require the ability to read. In former days, illiterate men could qualify for Army service, and many did. Under the new system, some functional illiterates are still getting through, that is, read below the fifth grade level, but the number has been reduced to an estimated five percent.

Based on the promise of more accurate measurement of job potential, the new ACB was adopted for operational use, and it was implemented on 1 May 1973. Now let's see how it is being used.

Operational use of the new ACB

The use of the new ACB has gone through two phases already in its short life. In the two months of May and June, a separate AFQT was administered to all Army male applicants. Beginning 1 July, experimental mental standards were adopted by the Army in which AFQT was dropped as a separate test, and an AFQT score is obtained from the ACB. The experimental mental standards were also changed on 1 July to require one aptitude area score of 90 or better for high school graduates and two scores of 90 or better for non-graduates, in addition to a percentile score of 10 or better on the AFQT obtained from the ACB.

The new aptitude areas made the experimental standards possible. Since each one contains some measure of general mental ability, plus specific aptitude measures, a qualifying score of 90 is more indicative of ability to succeed. The old aptitude areas did not measure enough general mental ability to use with the same degree of confidence. As of the time of writing no decision has been made about which standards to use as the official ones.

The ACB is now given at the time of application of enlistment, which provides the recruiter, counselor and applicant with the best possible information about the applicants mental qualifications and the job areas in which he is best qualified.

The tests are administered at AFEES or by Mobile Examining Teams. METs, who take the tests to the applicant. The entire battery requires about three hours to administer, which is longer than the previous testing time at AFEES, but the extra time is well spent. The new recruit actually spends less time in testing than before. He used to spend up to eight hours between AFEES and reception stations taking tests, but now he spends only the three hours at AFEES, plus another half-hour taking the

Auditory Perception, or Radio Code test, at AFEES if he is applying for training requiring radio code, or at reception stations, if he is not asking for a commitment for these jobs.

The new ACB is also given to women applicants at AFEES or by the Mobile Examining Teams. An AFWST, i.e. Armed Forces Womens Selection Test, score is also obtained from the ACB. No additional testing of women applicants to determine qualification is required. The new ACB is also being readied for use by reserve components and for retesting at posts, camps, and stations throughout the world. Soon the entire Army will be switched over to the new ACB. The Marines also are using the new ACB, but just as with the Army, the transition is not yet complete.

Evaluation of the new ACB.

One consideration in evaluating the new ACB was its effect on the flow of marginal men, or those with low general mental ability. Prior to introduction of the new battery, our research had shown that fewer men with low AFQT scores would obtain qualifying aptitude area scores of 90 or better. There was some concern that the effects would be serious enough to affect the manpower flow. Our argument was that the Army has little difficulty in recruiting men with low mental ability, and the experience since 1 May has borne us out. In July, when the ceiling on the number of men with marginal mental ability was removed, over a third of the accessions were in the marginal category. In August, the influx subsided, and the number was more acceptable at just over 20 percent. Army policy in the past few years has been to hold the number of marginal accessions to somewhat under 20 percent, and that is about the number coming in.

The quality of new accessions in the past was measured on the basis of AFQT scores, which provides a single index of general trainability. All men regardless of aptitude in specific job areas, were categorized on the basis of their general trainability. With the old aptitude areas, the AFQT did provide the best measure of quality and the procedure was appropriate. With the new aptitude areas, categorizing men on the basis of AFQT is no longer as appropriate. Since the new scores contain measures of general ability, the number of qualifying aptitude area scores at or above specified levels provides a more accurate description of the quality of the accessions.

We have been working with Army personnel management for some time to develop a new basis for categorizing quality. Although at the time of writing there has been no final resolution about the definitions of the categories, it is clear that the intent is to broaden the basis by including all the aptitude areas in developing an overall quality index. The basis that is finally adopted should make sense from all points of view; some considerations are that it should have a known relationship to the mobilization population, it should provide useful information to personnel management and it should not stigmatize the individual. One of the complaints about the AFQT mental categories is that the Cat IV's, or men in the marginal category, are generally assumed to be the misfits. While there may be no way to avoid stigmatizing completing, the new system should be descriptive of trainability in specific areas, which may help prevent a general labelling.

Another area of concern about tests is that of effectiveness of the tests for minority group personnel. The tests should be equally effective predictors of success for all groups, that is, blacks and whites with the

same test scores should have the same expected level of performance. The new ACB is equally valid for blacks and whites, and blacks and whites with the same test score have the same predicted level of performance in job training courses.

From another point of view, the question is: are blacks underrepresented in the Army accessions as compared to the population as a whole? And here the answer is no. In July, 1973, about one-third of the accessions were black, and in August, 20 percent. These numbers compare to about 20 percent in fiscal year 1973, and about 13 percent in population as a whole. The new ACB thus is not excluding an undue proportion of blacks.

The matter of racial bias in employment, in this case joining the Army, is of course more complicated than just a matter of test scores or even prediction of success from test scores. Test scores are only part of the system, and adjusting only the scores, without also adjusting the initial job training, utilization on the job, and advanced training, would hardly be a satisfactory solution for either the individual or the employer.

Another group that merits special consideration is women applicants. The new ACB is used to determine mental qualification of Army women applicants. They must qualify on the combination of Arithmetic Reasoning and Word Knowledge subtests, which provide the AFWST score, and in addition obtain at least two aptitude area scores of 90 or better. The Army has had little difficulty in meeting the quotas for WAC's. The ACB, however, was standardized on an all male sample, with both blacks and whites included, but no women. The question may arise whether the new ACB is fair to women. The answer is yes, to the best of our ability. Based on large samples of males and females in high schools, the population means on arithmetic and vocabulary, which constitute the new AFWST, are about the same for both sexes. As for the aptitude areas, both sexes receive the

same job training, and therefore women require the same aptitude as men. If the prerequisite for job training is 100 or average, on a particular aptitude area, women as well as men need to meet this prerequisite because students of both sexes need to master the same knowledge and tasks. The ACB scores for women are considered to be appropriate measures of their potential to perform satisfactorily in job training.

In summary, the new Army Classification Battery has been in use for about six months. It is administered at the time of application for enlistment to provide information about the applicant's qualifications in the different job area. The evidence is that it is not keeping out excessive numbers of marginal men, blacks, or women. Reports received from the field are generally favorable in that it is an efficient and effective method to assess the potential of applicants.

CHART 1

CONTENT OF NEW ARMY CLASSIFICATION BATTERY

GENERAL ABILITY TESTS

ARITHMETIC REASONING	(AR)
WORD KNOWLEDGE	(WK)
GENERAL INFORMATION	(GI)
MATHEMATICS KNOWLEDGE	(MK)
SCIENCE KNOWLEDGE	(SK)

PERCEPTUAL ABILITY TESTS

PATTERN ANALYSIS	(PA)
ATTENTION TO DETAIL	(AD)

MECHANICAL ABILITY

ELECTRONICS INFORMATION	(EI)
MECHANICAL COMPREHENSION	(MC)
AUTOMOTIVE INFORMATION	(AI)
TRADE INFORMATION	(TI)

INTEREST MEASURES

COMBAT SCALE	(CC)
ATTENTIVENESS SCALE	(CA)
ELECTRONICS SCALE	(CE)
MAINTENANCE SCALE	(CM)

NEW APTITUDE AREA COMPOSITES

Test		Aptitude Area Composites									
		CO	FA	EL	OF	SC	MM	GM	CL	ST	GT
General Ability Tests											
Arithmetic Reasoning	(AR)	AR	AR	AR		AR		AR	AR	AR	AR
General Information	(GI)		GI		GI						
Mathematics Knowledge	(MK)		MK				MK			MK	
Word Knowledge	(WK)					WK			WK		WK
Science Knowledge	(SK)							SK		SK	
Mechanical Ability Tests											
Trade Information	(TI)	TI		TI			TI				
Electronics Information	(EI)		EI	EI			EI				
Mechanical Comprehension	(MC)			MC		MC		MC			
Automotive Information	(AI)				AI		AI	AI			
Perceptual Ability											
Pattern Analysis	(PA)	PA				PA					
Attention to Detail	(AD)	AD							AD		
Auditory Perception	(AP)					AP					
Self Description											
Combat Scale	(CC)	CC									
Attentiveness Scale	(CA)		CA		CA				CA		
Electronics Scale	(CE)			CE							
Maintenance Scale	(CM)						CM				

THE DEVELOPMENT OF THE ENGLISH COMPREHENSION LEVEL SCREENING TESTS FOR FOREIGN STUDENTS

BY: Cortez Parks

The English Comprehension Level (ECL) testing system is the primary quality control used in the English Language Training Program (ELTP) conducted by the Defense Language Institute, English Language Branch (DLIEL) at Lackland Air Force Base and at numerous overseas locations.

It is a unified testing system in that all test forms are originated and controlled by one agency. The construction, validation, distribution, administration, and control of the various types of tests prescribed by DLI are the responsibility of the Tests and Measurement Branch, Development Division, English Language Branch.

Among the tests produced by the Tests and Measurement Branch are the English Comprehension Level Screening Tests, or "ECL" tests as we refer to them, which are the subject of this paper. The ECL Screening Tests consist of general English tests, usually called ECL tests and the Specialized English Terminology (SET) tests.

The ECL test is designed to measure an individual's English listening and reading comprehension and to ascertain his capability to acquire knowledge of and to function effectively in English language training and work situations.

The ECL test does not measure speaking or writing ability. These skills are measured subjectively by the students' instructors.

This is done by rating the speech or writing production by the use of a speaking and writing proficiency level code key. Correlation studies made at DLIEL of the instructors evaluation compared to the ECL indicate that the subjective evaluation of these skills has a high correlation with the ECL scores.

ECL tests, as well as the SET tests, have been correlated with language performance of native speakers of English in various specialized fields (Electronics, Maintenance, - - - - etc.).

The minimum ECL scores and cutoff scores of the SET tests have been established for different levels of training in the various specialties of the Army, Navy, and Air Force. These scores represent the minimum English comprehension level necessary for foreign students to successfully cope with English instruction in tasks involving varying degrees of difficulty and/or danger.

An individual cannot normally be expected to perform satisfactorily and efficiently when exposed to training in those fields with an ECL score below the minimum requirements.

A 100 ECL score indicates that the foreign students can use standard training materials without any difficulty.

A 0-25 ECL includes beginning language students with no previous English language background or those with a weak English background.

A 0-39 ECL indicates that the student is at the elementary level. Language instruction given at this stage consists principally of carefully selected basic vocabulary, sentence pattern and related pronunciation drills. No formal specialized instructions is attempted.

A 40-59 ECL indicates that the student is at the intermediate level.

A 60-69 ECL is the qualification level for direct entry into CONUS or 3rd country apprentice specialized training.

A 70-79 ECL is the required qualification level for direct entry into CONUS or 3rd country in most basic courses (e.g. primary pilot training, etc.).

80 and above is the qualification level for professional career and advanced courses (C & GS schools, surgeons and advanced medical courses, pilot transition, advanced flying safety officers. demolitions, control tower operators, etc.).

ECL levels are not percentage scores but are scores converted to established qualification levels; i.e. an individual with a 70 ECL knows much more than twice the "amount" of English than one with a 35 ECL, and the progressive proportions between 45 and 90 ECL are not the same as those between 35 and 70. The ECL spread "value" is not linear, but is more like a modified logarithmic curve.

The language proficiency levels required for entry into the various training courses are set by the services responsible for operating the training facilities. Applicants failing to meet the language proficiency level requirements for direct entry are programmed through the English Language Branch for additional language training.

In early 1962, a number of the service schools reported that a significant number of students were entering training courses who were meeting the ECL requirements but who were deficient in the specialized terminology peculiar to their specialties. This was especially true in

the areas of electronics, maintenance, and weather. The service schools were finding it necessary to give these students extra-curricular instruction in technical terminology. At a Tri-Service ECL Test Conference at Air Training Command (ATC) in 1964, a decision was made to add SET tests to the ECL screening tests.

SET tests were developed in five specialized areas; electronics, maintenance, supply, weather, and medical. Thus, Series 6500 was the first series of the ECL screening tests that consisted of ECL and SET tests.

By use of the SET tests, Military Assistance and Advisory Groups (MAAGs) were alerted to the necessity of programming students through the DLIEL Specialized phase prior to their enrollment in CONUS training courses.

As the result of a language training survey conducted by DLI upon its conception in 1966, it was decided to add General Army and General Navy terminology to the screening tests. These two additional tests were incorporated into the 6700 Series screening tests. Also, several of the technical training schools were requested to submit lists of technical terms along with basic course materials to DLIEL for use in the development of SET tests. These materials are updated periodically.

The 7200 Series SET tests now in use and the 7400 Series which will replace the 7200 Series in July 1974, consist of two forms each of Electronics, Maintenance, Supply, Weather, Medical Professional, Medical Service, General Army, and General Navy tests that measure the comprehension of technical terminology considered essential by the respective technical training schools.

The ECL test consists of 120 questions, 60% of which are on tape and test the aural comprehension of the student. 40% of the questions test reading comprehension. 40% of all objectives tested are taken from the ALC Elementary and 60% from the Intermediate texts. Items are presented in the form of questions, statements, and dialog in the listening part of the test and in the form of incomplete statements, underlined objectives, word order, and complicated sentences in the written part. 75% of the questions consist of vocabulary and idioms and 25% test the students' abilities to select the proper structure. All items are of the multiple choice type with four choices of answers, only one of which is the correct answer. The standards for the construction of items are extremely rigid.

The resulting product is the service trial edition of the screening test consisting of 12 forms of ECL tests and two forms of each of the SET tests.

Validation of the ECL and SET tests are accomplished by pretesting the service trial edition with the foreign students at the various technical schools. Since the students at the technical schools are all in the upper quadrant of language proficiency, the ECL tests are pretested also among students in the lower quadrant at overseas schools.

The number of cases used for experimental testing depends upon the type of statistical analysis required. For conducting item analyses, at least 100 cases were used. For computing correlation or for norming, at least 300 cases were obtained.

A great amount of effort is being expended in the development of the ECL screening tests in order to obtain the highest reliability and validity for each form of the tests. For test reliability, both the internal item consistency and the alternate-form reliability are carefully checked. For test validity, we check the content validity, construct validity, and predictive validity. The reliability index and the construct validity coefficient of each of the different forms have been kept above .90. The coefficients of the predictive validity have shown above .65. All of these cutoffs are considered high standards in test validation. For an example of statistical data, please refer to handout, Table I.

We have also correlated the ECL tests against the Test of English As A Foreign Language (TOEFL) tests, the University of Michigan language proficiency tests, and the American University language tests. Pretesting has also been conducted at San Francisco State College, the University of Texas, Texas A&M, and the University of Minnesota. The results were around .80, indicating a high correlation with their criterion tests of English proficiency.

The SET tests do not use converted standard scores as do the ECL tests. They are designed to use percentage scores with pass/fail cutoff scores. For the 6500, 6700, 6900, 7200, and 7400 SET tests, we have used minus one standard deviation as the cutoff point since these are proficiency type tests.

TABLE I

	EL-69		MA-69		SU-69		WX-69		ME-69		AR-69		NA-69	
	A	B	A	B	A	B	A	B	A	B	A	B	A	B
Maximum possible score	100	100	120	120	100	100	100	100	120	120	100	100	100	100
Mean Score	70.4	70.9	81.2	77.0	54.1	53.6	62.8	63.3	87.3	87.8	66.5	67.0	63.5	63.3
Standard Deviation	15.8	15.3	25.6	24.2	18.1	17.8	20.3	20.1	24.1	24.9	15.0	13.7	15.5	15.3
Mean Difficulty Index	.65	.67	.62	.59	.57	.57	.65	.63	.64	.65	.66	.66	.62	.63
Mean Discrimination Index	.43	.37	.49	.48	.47	.47	.49	.50	.43	.44	.37	.32	.37	.36
Reliability Index for Internal Consistency (Kuder-Richardson Formula 21)	.93	.92	.97	.96	.93	.93	.95	.95	.97	.97	.91	.89	.91	.90
Alternate-form Reliability Coefficient (Pearson Product-Moment Correlation Formula)	AxB .87		AxB .96		AxB .89		AxB .95		AxB .94		AxB .88		AxB .92	
Validity Coefficient of Correlation between Specialized Terminology Tests and ECL Tests.	.79	.79	.88	.86	.77	.75	.86	.86	.76	.76	.70	.72	.61	.69

AN EXPERIMENTAL, MULTIMEDIA CAREER DEVELOPMENT COURSE FOR NEW MENTAL STANDARDS AIRMEN

George P. Scharf

Air Training Command - USAFSAAS, Chanute

This paper is a report on an experiment to determine if modifying the Fire Protection Specialist Career Development Course (CDC) could improve the CDC as a training device for low mental aptitude airmen. Modifications to the regular CDC included simplifying the written materials and adding 18 hours of audio supplementation. The experiment used three groups of New Mental Standards (NMS) Project 100,000 airmen enrolled in three versions of the CDC. The control group took the regular fire protection specialist CDC. The first experimental group took the modified CDC with audio supplementation. The second experimental group took the modified CDC without any audio supplementation. Criterion for effectiveness was the final grade performance of the personnel in the three groups. Results indicated both experimental groups had mean final grade scores significantly better than the control group. However, there was no significant difference between the mean final grade scores of the two experimental groups. Mean course completion times for the three groups were within a week of one another. In this experiment, the audio-supplemented CDC was not a cost-effective method of course presentation.

In recent years the federal government has been quite concerned with improving the utilization of the nation's manpower resources. In the late 60s and until 1971 there was considerable pressure on the military services to accept and train personnel with low mental abilities. One such program, Project 100,000, began in 1966. This program called for the military services to accept up to 100,000 enlisted personnel scoring between the 10th and 30th percentiles on the Armed Forces Qualification Test (AFQT). These are persons in the Category IV mental group.

As a part of Project 100,000 within the Air Force, a number of career fields were selected for implementing the use of large numbers of marginal aptitude, New Mental Standards (NMS) personnel. While NMS airmen were a part of Project 100,000, not all Project 100,000 airmen were NMS people. The definition of New Mental Standards airmen is, "those who have scores in the lower half of the Mental Group IV range." One of the career fields selected to utilize NMS airmen was that of Fire Protection Specialist, AFSC 571X0.

NMS personnel were assigned to the Fire Protection Specialist Course at Chanute directly from Lackland. However, to merit this assignment, the

student had to have (or achieve before leaving Lackland) a reading level of at least the sixth grade. When the NMS airmen entered training, it was found that most of these personnel with low mental abilities and low reading levels could complete the Fire Protection Course if given adequate remedial instruction, counseling, and personal guidance. On the job, however, where structured remedial and personal help were not as readily available, low reading ability caused great difficulty for these airmen in studying and completing their Fire Protection Specialist Career Development Course (CDC). As a result, ATC Headquarters directed that an experiment be set up to determine if simplifying the CDC and adding audio supplementation would enable students to complete their CDC faster and with better grades.

The experimental design finally selected called for comparing three equivalent groups of NMS airmen who had successfully graduated from the basic Fire Protection Specialist Course at the School of Applied Aerospace Sciences, Chanute. The AFQT scores for the NMS experimental subjects ranged from 20 down to 10. Upon reaching the field and applying for their CDCs, these airmen were enrolled in one of three different versions of the Fire Protection Specialist CDC. The students in all three groups took the same pre-test and final or post-test. The three CDC versions covered the same basic information, but their modes of presentation differed as follows:

CDC 57151A. Airmen who took this course were the control group. This is the conventional Fire Protection Specialist CDC.

CDC 57100. This was an experimental, multimedia CDC which used short sentences, large, single-column-per-page print, contained more pictorial materials than the regular CDC, and used cassette tape recordings to summarize, emphasize, and express in different terms the information read by the student. No new or different information was included in the recordings. While detailed instructions on using the tape player with the written material are a part of the CDC text, the procedure was somewhat as follows:

(1) After reading an instruction sheet on the tape player and directions at the start of the first CDC Volume, the student studied the first part of his CDC. After a few paragraphs, he was instructed to put a particular cassette into the tape player and turn it on.

(2) The tape recording referenced the passages the student had just read. Speaking in turn, a group of up to four male voices then discussed, summarized, restated, or reviewed the information in the referenced paragraphs. The student was then directed by a recorded voice to turn the player off. He then returned to the written text until he reached the printed instruction to again turn on the tape player. This procedure was followed throughout the entire five volumes of the CDC for a total of approximately 18 hours of audio review of what the student had read.

(3) Volume introductions, chapter introductions, and chapter summaries were recorded with a female voice to give more variety to the tapes.

CDC 57100B. This was the same, less verbal, more pictorial CDC 57100 without the tape recordings and with all references to using a tape player removed from the written text.

Procedures

Starting in August 1971, a graduation roster of each class in the Fire Protection Specialist Course was furnished the Training Research Applications Branch. The first entries into the CDC experimental program graduated from the Chanute course on 7 September 1971. When an NMS airman graduated, his name and social security account number were furnished to the Extension Course Institute, Educational Systems Branch (ECI/EDXV), at Gunter AFB in Alabama. These names were then put on an identifier list at ECI as eligible for the CDC experimental program.

a. The number and flow of NMS personnel available for this experiment never became as great as the originally projected 16 per month. One reason for this was that Project 100,000 unofficially decreased to Project 50,000 some months before the experiment began. As a result, the number of NMS students entering into the Fire Protection Specialist Course decreased by more than half. Thus, by the time the CDC course materials were prepared, there was a shortage of eligible participants for the experiment.

b. Effective 1 April 1972, Change 4 to AFM 33-3, Enlistment in the Regular Air Force, cut off the NMS student flow entirely. Under this change, Air Force enlistees must have an AFQT of at least 21, and if the AFQT is between 21 and 30 the enlistee must be at least a high school graduate. This change in enlistment requirements stopped entry into the Air Force (and thus eventually into this experiment) of NMS airmen. As a result, the number of cases analyzed in the study is as follows: regular CDC -- 24, experimental group 1 using the tape recorder -- 25, experimental group 2 using only the rewritten course materials -- 19.

As applications for the Fire Protection Specialist CDC were received by ECI, the NMS identified personnel were assigned to one of the three groups described above, i.e., CDC 57151A (control group), CDC 57100 (multimedia experimental group), CDC 57100B (second experimental group).

Before being sent his course materials, each student was administered a 100-question pre-test by his OJT supervisor. When ECI received the pre-test, the written CDC materials for the course to which he was assigned were sent to the student. The tape players for the multimedia CDC 57100 were sent from Chanute. Students in all three CDCs took the same pre-test and final test.

The measures of effectiveness of the two experimental CDC groups was an analysis of their final course grade performance in comparison to the control CDC group.

All calculations were performed using statistical packages available in the PLATO IV computer-based teaching system. There are presently four PLATO IV terminals at Chanute connected to a central computer located at the Urbana campus of the University of Illinois. The mathematical formulas necessary for any analysis, including those involving unequal n's, are included in the analysis programs used. The person making the analysis is required only to select the type of analysis desired and input the raw data. Answers appear on the screen immediately. Thus, it was possible to examine a great amount of raw data concerning the subjects of the study and receive back an almost instantaneous readout of the analysis selected.

Findings

Overall results obtained from this study are summarized in Table 1.

Table 1

	CDC 57151A	CDC 57100	CDC 57100B
	n=24	n=25	n=19
AFQT mean	15.33	15.08	15.05
AFQT standard deviation	3.41	3.55	3.40
Resident course grade mean	83.75	83.88	84.61
Resident course grade standard deviation	3.43	3.69	3.68
Pre-test grade mean	56.58	57.40	61.95
Pre-test grade standard deviation	10.58	9.57	10.78
Post-test grade mean	66.75	72.80	76.00
Post-test standard deviation	7.71	10.73	11.07
Net-mean grade score increase	10.17	15.40	14.05
Net-mean grade score increase standard deviation	9.59	8.72	8.98
Completion time mean	205.00 days	211.50 days	205.32 days
Completion time standard deviation	76.45 days	50.26 days	31.64 days

To establish that the three groups were equivalent in ability, a one-way analysis of variance was conducted on the means of the students' AFQTs, final grades from the resident Fire Protection Specialist Course, and their CDC pre-test scores. Table 2 summarizes the one-way analysis of variance on each factor among the three groups.

Table 2

Source of Variation	F	d.f.	p
AFQT	0.045	2,65 =	0.9558
Resident Course Final Grade	0.341	2,65 =	0.7122
CDC Pre-test Grade	1.628	2,65 =	0.2042

As can be seen, the F ratios for each of these three student variables were all non-significant, thus indicating comparable groups.

In order to test the effectiveness of the instructional methods, a one-way analysis of variance was conducted on the means of the post-test CDC (final grade) scores achieved by the three groups. Table 3 presents this analysis of variance.

Table 3

Source of Variation	F	d.f.	p
Post-test score	4.966	2,65 =	0.0098

Inspection of Table 3 shows that the post-test score F ratio was 4.966; this is a significant value indicating the post-test performance of the three groups was not comparable.

To compare the differences of the three groups on the post-test performance factor, a series of t tests were run. Their results are summarized in Table 4.

Table 4

T_1	T_2	t	d.f.	p
Mean post-test score of Ex Gp 1 (revised text, audio supplement)	Mean post-test score of control group			
72.80	66.75	2.258	47.0	.06
Mean post-test score of Ex Gp 2 (revised text only)	Mean post-test score of control group			
76.00	66.75	3.227	41.0	.05
Mean post-test score of Ex Gp 2 (revised text only)	Mean post-test score of Ex Gp 1 (revised text, audio supplement)			
76.00	72.80	.967	42.0	.44

The inferences from Table 4 are that both the audio supplemented CDC and the revised text only CDC yielded significantly higher grades than the regular CDC. However, there was no significant difference in the final course grades of the two experimental groups.

To establish the equivalency of the completion times, a one-way analysis of variance was made on their means. This analysis is shown in Table 5.

Table 5

Source of Variation	F	d.f.	p
Completion time of CDC in days	0.064	2,65	= .9383

The indication of this analysis is that the CDC completion times for the three groups were not significantly different.

Discussion

In this study, both experimental CDC groups had statistically significant better learning scores than the regular CDC students. However, there was no significant difference in the learning scores of the two experimental groups. While not statistically significant, the experimental group without the audio supplementation had a mean final grade that was 3.2 points higher than the experimental group using audio supplementation. At the same time, the net-mean grade score increase of these two groups was 1.78 higher for those using audio supplementation. Both experimental groups had greater net-mean grade score increases than the control group.

While costs of a program such as this are a rather elusive thing and hard to pinpoint, it would appear that the production, implementation, and administration of the audio-supplemented CDC was between three and four times the cost of the regular CDC. This factor remains about the same when projected over a much larger number of students than participated in this experiment. For all three CDCs the production of the written materials is equivalent. In addition, the larger the printing of a CDC, the less cost per student. However, the premise that the larger the audio group the less the cost per student does not appear valid. While there might be some slight savings in buying a large number of cassettes and tape players, the loss rate plus maintenance and replacement costs of these items would nullify any projected reductions in cost per student based on a large number of users. Consequently, it does not appear cost effective for widespread adoption of an audio-supplemented CDC program when students did no better (on the average) than the students who received only the revised text. Sellman, 1970, in a similar experiment used subjects of low, medium, and high aptitude and found that it was low-aptitude personnel who showed the greatest learning gains when audio supplementation was used. Since, at this writing, these low-aptitude personnel are not accepted by the Air Force, the use of audio supplementation to CDCs would appear to have even less potential as a cost effective teaching device in the CDC program.

Of the 31 NMS airmen assigned to the 57100 course, four were discharged before completing training, one desserted, and one is in confinement. This leaves the number of NMS subjects who completed training at 25. Nine tape players were lost or stolen.

At the end of CDC 57100, students were asked to comment on the use of the tape player and its effect on the course. In general, the reaction of students was favorable, but not overwhelmingly in support of the audio tapes. The student's supervisor was also asked to comment on the use of the tape player. Here too, the general reaction to the tape player was favorable with one or two of the supervisors very strong in their support of the idea and one or two as strongly opposed. Fifteen of the supervisors seemed to feel the tape helped the student to finish faster than normal.

However, the mean time for those using the audio tapes was a week longer than those using the other versions of the CDC. The difference in completion times for the three groups was not significant.

Conclusions

As a result of the analysis of the data evolved from this study, the following conclusions are hereby stated.

The two experimental groups had better final CDC grades than did the control group.

The differences in grades of the two experimental groups was not statistically significant. The audio-supplemented experimental group had a net-mean increase of their final grade over their pre-test grade 1.78 points more than the non-audio supplemented group; but the non-audio-supplemented group had a mean final grade 3.2 points higher than the audio-supplemented group.

Any higher grades of the experimental groups can be attributed to the revised text as much as to any audio supplementation as all three groups were shown as equivalent in mental ability and pre-course knowledge.

Mean completion times for the three groups were within a week of one another.

Audio supplementation was not a cost-effective means of instruction in this experiment.

References

Effectiveness of Experimental Training Materials for Low Ability Airmen, Waynce S. Sellman, Captain, USAF, AFHRL-TR-70-16, June 1970.

Optimal Utilization of On-the-Job-Training and Technical Training School

by

Captain Alan D. Dunham

This paper discusses work planned and completed during the AF Human Resources Laboratory's effort to develop techniques for optimizing the use of On-the-Job Training (OJT) and Technical Training School (TTS) as methods of initial upgrading for non-prior service personnel.

I. INTRODUCTION

The author's previous MTA paper (1972) discussed criteria relevant to selecting optimal OJT/TTS mixes and then described the results of work which established the feasibility of obtaining OJT cost data.

Significant improvements in the OJT costing methodology have been obtained in the year since the first report and progress has been made in the application of a measure of one other relevant criterion, the quality of the training methods. This progress is discussed in the following text.

II. REFINEMENT OF OJT COSTING METHODOLOGY

The original OJT costing methodology[1] identified the following cost factors:

Table 1

OJT Cost Factors

Student Time
Instructor Time
Delayed Entry into Training
Records Management
Remedial Training
Equipment and Materials

1 Dunham, 1972

Data concerning these cost factors were collected by survey for the Communications Center Operations specialty for OJT to the 3 (semi-skilled) level for non-prior service airmen. These surveys asked supervisors in this job specialty to provide detailed data on several items, including time spent learning a number of skills specific to Communications Center Operations. The Supervisors had a wide range of experience with OJT. The resulting OJT estimate was considerably less than the cost of TTS for the corresponding training course.

The most important aspect of this research was that it established the feasibility of obtaining cost estimates for USAF OJT through the use of a survey technique.

One undesirable result was the high response variance encountered for many of the crucial questions. This resulted in relatively wide confidence limits for the OJT cost estimate, but not wide enough to cast doubt on the cost comparison. Another undesirable characteristic was that a large section of the survey instrument was specific to Communications Center Operations. A survey instrument independent of the job performed would require less resources for survey administration and data file maintenance.

III. THE DEVELOPMENT AND EVALUATION OF ALTERNATIVE METHODOLOGIES FOR ESTIMATING THE COST OF AIR FORCE OJT

In June of 1972 Manpower and Personnel Systems Division of the AF Human Resources Laboratory let contract #F41609-72-C-0048, designed to improve upon the original OJT costing methodology. During the first phase of this contract the contractor used the original survey approach and two alternative survey techniques to simultaneously collect cost data concerning OJT to the 3-level for USAF Administrative Specialists.

One of the two new alternative survey techniques asked supervisors to estimate information for eleven variables for the previous week, including time spent on OJT by instructors and trainees. The other alternative survey technique required supervisors to record time spent by instructors and supervisors on a daily basis for one week.

Under ideal circumstances each survey technique would be used on a unique set of supervisors. The time-phasing of the contract and an unexpected reduction in the number of Administrative Specialists entered in OJT forced administration of all three survey instruments to the same set of supervisors. The three distinct survey techniques were preserved

in the instrument, but most of the redundant questions were asked only once to minimize the length of the survey.

Of the 295 surveys sent to supervisors at 88 Air Force bases, 207 were returned. All responses were manually edited and keypunched for transferal of the survey data to magnetic tape.

The questionnaire's length might have been responsible for a reduced response rate toward the end of the instrument, but response rates for all three survey approaches were sufficient to enable comparison of the alternative techniques.

A detailed report of the survey data and an in-depth discussion of their implications will be forthcoming soon in an AF Human Resources Laboratory technical report. For brevity and to minimize duplication Table 2 describes item content and mean responses without discussion of their implications. The reader is cautioned against inferring strong conclusions from this limited display of data.

A cost estimate for each OJT cost factor is exhibited in Table 3. The cost models used to compute these estimates were selected from several cost models based on various combinations of the three survey approaches. Again, a more rigorous discussion will soon appear in AFHRL publication. The OJT cost per trainee of $1545 can be compared to the marginal cost per trainee of course 3ABR 70230, $2271.[2] There are criteria other than costs which are equally important, viz., training capacities and quality of the training. Thus, conclusions regarding optimal OJT/TTS mixes for the Administrative Specialty based only on data in this report would be reached without considering other important criteria. These criteria are discussed in a Technical Report in review at AFHRL.

2 The cost of Tech school was obtained from RAND Corporation data.

Table 2

Mean Responses for Admin Specialist OJT Survey
Responses Based on Supervisors' Total Experience

Item Content	Mean Response
Persons in Section	7.27
Days Delay in start of 3-level OJT	8.60
Days Delay in start of 5-level OJT	8.44
Weeks to 3-level Proficiency	13.30
Weeks from Proficiency to Actual Award of 3-level	3.31
Number of 3-level Trainees	1.34
Number of 5-level Trainees	1.06
Additional Training Capacity	1.11
Operate with fewer personnel if no OJT	29% [3]
Are Trainees Productive?	82% [3]
OJT Superior to TTS Grad	22% [3]
TTS Superior to OJT	32% [3]
No Difference between OJT and TTS Grads.	46% [3]
Percent Failing Apprentice Knowledge Test	25.64
Weeks Remedial Training	3.60
Trainee Remedial Hours/Week	9.07
Instructor Remedial Hours/Week	5.28
Training Record Keeping Hours/Week	1.94

Skill Area[4]	Wks. to Prof.	Instructor Hrs/Wk.	Trainee Hrs/Wk.
Career	2.86	2.94	7.78
Security	2.05	1.78	3.89
Supervision	4.80	3.22	6.90
Equipment	3.78	3.31	9.14
Publications	4.51	3.59	7.83
Forms	3.64	2.65	5.43
Communications	4.99	3.27	8.28
Documentation	5.02	3.04	6.92
Library	2.70	1.16	4.02
Postal	3.10	2.07	5.04

3 Per cent of total valid responses that responded affirmatively.
4 Corresponds to the Job Proficiency Guide skill breakdown for this speciality.

Table 2 (Cont'd)

Week	Productive Trainee Hours (Mean Response)
1	11.87
4	17.92
8	23.44
12	27.25
16	30.78
20	33.21

Responses Based on Supervisor's Experience Previous Week:

Item Content	Mean Response
Week of Training for Trainee	10.88
Percent of Training Completed	56.21
Percent of Skills known prior to OJT	10.93
Days of delay in start of 3-level OJT	9.16
Trainee Hours OJT	9.07
Instructor Hours	10.88
Record Keeping Hours	1.88
Days of delay in start of 5-level OJT	6.71
Percent of skills known at time of arrival at unit (TTS Grads)	29.57
Additional weeks to 3-level Proficiency	5.05

Responses Based on Supervisor's Record-Keeping for one Week:

Item Content	Mean Response
Number of Trainees	1.3
Trainee hours instruction	10.61
Trainee Hours productive	24.73
Instructor hours/trainee	8.80

Table 3

Summary of OJT Costs

Cost Factor	Cost/Trainee
Trainee Time	$ 570.18
Instructor Time	591.35
Delayed Entry	196.05
Records Management	137.70
Remedial Training	40.03
Equipment and Material	10.18
	$1545.49

Tables 4 through 6 summarize the more obvious conclusions concerning the three survey approaches used. When sample size is no problem and when information specific to the skill groups is not needed, the survey approach used should be that of Table 6. When sample size is a problem, the methodology of Table 5 would suffer from the same problem as that of Table 6, so that a form of the original survey approach appears more advantageous. Most limitations on sample size can be overcome by extending the time span of the survey.

Selection of a survey approach thus involves trade-offs. The Phase I report of this contract recommends use of the general survey approach of Table 6 while retaining some of the questions used in the original methodology. This synthesis of the survey approaches was used in Phase II of the contract and appears as Appendix A of this paper.

Table 4

Original Survey Approach - Responses Based on Supervisor's Total Experience

Advantages (compared to alternative methodologies):

a. Larger sample at any point in time

b. Breaks out information by skill groups specific to speciality or job performed

c. Can be completed in one sitting - minimizing turnaround and survey control workload

Disadvantages:

a. Speciality - dependent

1. makes aggregation of OJT cost data across specialities more complex

2. requires that a different survey form be developed for each speciality

3. requires multiple formats for maintenance of OJT cost data.

b. High response variation, resulting in relatively high variance around OJT cost estimate.

Table 5

Alternative Survey Approach - Responses Based on Supervisor's Experience During Previous Week

Advantages:

a. Can be completed in one sitting;

b. Speciality-Independent;

c. Variance reduced somewhat compared to original methodology;

d. Recall required is for more recent experience.

Disadvantages:

a. Provides no data on costs related to specific skill groups;

b. Requires presence of OJT trainee during previous week, which may result in a reduced sample size for any point in time.

Table 6

Alternative Survey Approach - Responses Based on Record-Keeping for One Week

Advantages:

a. Generally reduced response variance compared to other survey approaches;

b. Speciality - independent;

c. Recall involved limited to one day of work.

Disadvantages:

a. May have reduced sample size due to need for currently enrolled trainee;

b. Turnaround time is longer;

   1. may result in reduced sample size due to increase in time in the field;

   2. will require more record keeping by survey control.

Phase II of the OJT costing contract is nearly completed. During this phase the contractor utilized the survey technique developed in Phase I to collect cost data for OJT to the 3-level for the following specialities: Fire Protection, Pavements Maintenance, Fuels, Materiel Facilities, and look.

The results of this phase will soon be ready for publication. Persons interested in these results should be able to obtain information from Manpower and Personnel Systems Division of the AFHRL in the near future.

IV. QUALITY OF THE TRAINING METHODS

The OJT costing work described above shows that the Air Force can and is obtaining OJT cost estimates compatible for comparison with TTS cost estimates. One assumption required for a straightforward comparison is that these two methods of training provide trained personnel of equal quality in terms of their productivity on the job. Cost estimates which included 'quality' in some manner would be tenuous figures based on unsupported assumptions because measures of quality useful for this purpose do not exist. Hence, cost estimation and quality comparisons are treated separately.

The following discussion of methods of quality comparison is a summary of some work completed at Manpower and Personnel Systems Division, AFHRL. A thorough analysis of the 'quality' problem is in preparation and should appear as an AFHRL technical report within a year.

The problem of comparing the quality of training methods in the military context deserves a thorough theoretical discussion of the impediments faced by researchers. It was thought expedient to publish some interesting data in this article with only a short discussion of problem traits to allow interested persons an earlier glimpse of the results.

Previous research concerning this subject is limited. One exercise was the author's first simple comparison of mean Speciality Knowledge Test (SKT)[5] scores, which essentially showed no differences between the performance of OJT and TTS graduates on the SKT for the Communications Center Operations specialty. This may have been indicative of future

5 Dunham, 1972

results but lacked depth in that it did not account for related variables that may have some concomitant effect on SKT scores, such as the pre-training abilities of the trainees.

Another AFHRL study, The Road to Work: Technical School Training or Directed Duty Assignment?, by Mr. W. B. Lecznar, used several criteria which are thought to be related to performance, including: Job Difficulty Composite[6], Number of Tasks Performed (range 1-372)[7], Average Task Difficulty per Unit Time (range 1-7)[7], and operationally prepared Airman Performance Report (range 1-9) overall evaluations. Several variables thought to cause variation in these criteria were used in addition to OJT/TTS as grouping in an analysis of covariance or, "...a two treatment-one concomitant variable multiple regression model."[8] Analyzing these results, Mr. Lecznar concluded that, "The inclination is to say that in these specialties formal resident technical training provides little or no advantage over on-the-job training."[9]

Airman Proficiency Ratings (APR's) would appear to be a 'natural' for measuring some aspects of performance, but they currently suffer from such a high degree of inflation that their lack of variation severely limits their potential for this purpose.

The Navy has also done some interesting work in developing productivity indices.[10] This concept may eventually be a fruitful avenue for future work but a few shortcomings must be overcome. One problem is that it will be difficult to account for changes in the quality of output or in the type of output. In addition, output level in many work centers in the Air Force is strongly related to the demand for the output, which could be a troublesome variable to account for in a productivity index.

A logical attack for this problem would be to select the best measure of productivity, performance, or proficiency, a priori, and then accept the results obtained with observations on that measurement. State-of-the-art for this type of measurement is

6 Meade & Christal, 1970
7 Lecznar, 1972, p. 2
8 Ibid, p. 3
9 Op. Cit., p. 9
10 BUNAVPERS, 1969

still primitive, so measurements used are those available, which are usually designed for another purpose. This is true of Specialty Knowledge Tests (SKT's).

SKT's are used operationally by the Air Force as promotion criteria. They are specialty-specific paper and pencil tests updated frequently by Air Force supervisors with current experience in the specialty. An SKT score for an individual may be said to indicate that individual's potential for performance or productivity within the given specialty. SKT's are taken for promotion purposes, thus all with a score for a particular SKT had a common motive for excelling on the test.

If either OJT or TTS were 'superior' to the other, one would expect this to be manifested in the graduates' performance. Studies discussed above found no evidence to support the hypothesis that OJT and TTS differ in the quality of their graduates. Because SKT's are specialty-specific, used by the Air Force for promotion, and are taken by both OJT and TTS graduates with common motives for excelling, SKT's may be useful for examining this hypothesis.

Thus, SKT scores were used to compare the specialty knowledge of OJT and TTS graduates who received initial upgrade training in the same specialties for which OJT cost data has been or is being collected. The research hypothesis tested was that OJT and TTS graduates are of equal quality. Operational testing of this hypothesis was conducted by comparing the individual SKT scores of OJT graduates with those of TTS graduates, holding constant other factors thought to be related to SKT scores. The rationale for including or excluding 'related' variables will not be developed here. Analysis of covariance was used to test the operational hypothesis and also to test whether the 'related' variables explained some of the variance of the SKT scores. Table 7 lists the variables included for each test specialty. Separate analyses were conducted for the specialties listed in Table 8.

Table 7

Variables Used in the Analysis of Covariance

Variables	Description
SKT Percent Right	Individual's first score for Test Specialty
TESTIM	Days between date of enlistment and date of test administration
NEGRO	Categorical: 1 if Negro, Ø if other
FEMALE	Categorical: 1 if Female, Ø if other
ADMINISTRATIVE AI GENERAL AI MECHANICAL AI	Aptitude indices as computed from the Airman Qualifying Exams taken prior to enlistment. Values are in raw score form.
TEST REVISION NUMBER	SKT's are systematically updated. This is a set of categorical variables indicating which test form an individual received a score for.
OJT	Categorical variable indicating method of initial training. 1 of OJT, Ø if TTS
OJT X ADMIN AI	Interaction between OJT variable and Admin AI score
TTS X ADMIN AI	Interaction between TTS variable (1 if TTS, Ø if other) and Admin AI score
OJT X GEN AI	Interaction as above
TTS X GEN AI	Interaction as above
OJT X MECH AI	Interaction as above
TTS X MECH AI	Interaction as above

Table 8

Air Force Specialties Analyzed

Air Force Specialty Code[11]	Description
551X0	Pavement Maintenance
571X0	Fire Protection
622X0	Cook
631X0	Fuels, petroleum
647X0	Materiel facilities
702X0	Administrative Specialist

Table 9

Summary of Hypotheses

Null Hypothesis	Variables Tested
1	OJT x (appropriate) AI TTS x(appropriate) AI
2	OJT
3	OJT x (appropriate) AI TTS x (appropriate) AI OJT

11 Also, Test Specialty. The 'X' in each code is replaced by a 4 or 5 depending on whether the SKT was for promotion to E4 or E5 in that specialty. Both of the 4 and the 5-level SKT's were used for all specialties in the Table.

For each test specialty three regressions were computed in the following order and format:

1. $SKT = B_1(OJT) + B_2(NEGRO) + B_3(FEMALE) + B_4(OJTxAI) + B_5(TTSxAI) + B_6(TESTIM) + B_7(TEST\ FORM)$

2. $SKT = B_1(OJT) + B_2(NEGRO) + B_3(FEMALE) + B_6(TESTIM) + B_7(TEST\ FORM) + B_8(AI)$

3. $SKT = B_2(NEGRO) + B_3(FEMALE) + B_6(TESTIM) + B_7(TEST\ FORM) + B_8(AI)$

Three F tests were computed based on the increase in error sums of squares between, respectively, regressions 1 and 2, 2 and 3, and 1 and 3.

Null hypothesis number 1 concerned whether inclusion of the two interaction variables, 'OJT x AI' and 'TTS x AI', resulted in significant reduction in the error sum of squares. These variables provide for the possibility that ability, as measured by an appropriate Aptitude Index, has a differential impact on SKT scores obtained by OJT and TTS graduates. This is also sometimes called a 'slope' test.[12]

The second hypothesis concerned whether the OJT dummy variable significantly improved the error sum of squares. The regression did include an appropriate Aptitude Index score, but not in interaction form. This is the 'intercept' test. Rejection of the null hypothesis would support a contention that OJT and TTS graduates have different average SKT scores, holding other variables constant.

The final null hypothesis provided an 'overall' test. It tested whether inclusion of both the interaction variables and the OJT variable resulted in a significant reduction in the error sum of squares.

---

12 The technique used is computationally similar to that discussed in Lecznar, 1972. Wonnacott & Wonnacott, pp 77-79 and Johnston, pp 192-207, contain discussions of analysis of covariance using multiple linear regression as a computational approach.

AD-A040 773 AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TEX F/G 5/9
PROCEEDINGS, ANNUAL CONFERENCE OF THE MILITARY TESTING ASSOCIAT--ETC(U)
NOV 73

UNCLASSIFIED NL

2 OF 9
ADA
040773



Table 10

F Ratios for Test Specialties for Which the Reductions in Error Sums of Squares Due to OJT/TTS Variables were not Statistically Significant[13]

Test AFSC	H_o	DF_{NUM}	DF_{DENOM}	F_{RATIO}
55140	1	1	121	2.42
	2	1	122	3.30
	3	2	121	2.88
55150	1	1	1494	0.07
	2	1	1495	0.65
	3	2	1494	0.38
57140	1	1	2710	0.27
	2	1	2711	0.89
	3	2	2710	0.57
57150[13]	1	1	1914	0.07
	2	1	1915	4.37
	3	2	1914	2.22
62240	1	1	621	2.46
	2	1	622	1.32
	3	2	621	1.89
62250[14]	1	1	2982	4.19
	2	1	2983	5.77
	3	2	2982	4.98
63150[14]	1	1	1428	1.30
	2	1	1429	4.03
	3	2	1428	2.67
64750[2]	1	1	3909	5.55
	2	1	3910	2.38
	3	2	3909	3.97
70250[2]	1	1	6458	0.61
	2	1	6459	0.28
	3	2	6458	0.44

13 At the .01 level

14 TESTIM was included in these regressions. This variable was excluded from other F tests because it was not statistically significant at the .01 level.

Table 11

Test Specialties for which OJT/TTS Variables Proved to be Statistically Significant[15]

Test Specialty	OJT	NEGRO	FEMALE	GEN AI	MECH AI	TEST FORM	OJT X ADMIN AI	TTS X ADMIN AI	CONSTANT	R^2	DF N	DF D	F RATIO
63140	-1.27[16]	-2.29	4.40	-	.20	-6.36	-	-	39.42	.179	1	2509	26.66
64740	-2.29[16]	-1.71	-	.39	-	4.66	-	-	31.41	.246	1	2411	30.56
70240	2.63	-1.27	-0.72	-	-	-5.54	.19[16]	.24[16]	28.47	.203	1	10716	14.80

NOTE: Numbers in columns under variable names are regression coefficients.

[15] at the .01 level

[16] Variables for which the F Ratio was computed. Data used to compute these regressions and test statistics were obtained from files maintained at Personnel Research Division, AFHRL. Records, or observations, used were for personnel who, up to the time they took on SKT, served only in the specialty for which that SKT was written. Final computations were completed under Project 6323, Task 0209, Study 4903.

Table 9 summarizes these hypotheses. Table 10 lists the test specialties for which the F ratios would not allow rejection of the null hypotheses. For these test specialties no differences were found between the SKT scores of OJT graduates and those of TTS graduates. Table 11 displays regression and correlation statistics for the test specialties for which one or more of the null hypothesis was technically rejected. Figures 1, 2, and 3 show the data of Table 11 graphically, holding constant the dummy variables NEGRO, FEMALE, and TEST FORM.

Figure 1 illustrates the results for SKT test specialty 63140, Fuels Specialist. Acceptance of the second hypothesis means that there was a statistically significant difference in the performance of OJT and TTS graduates on the 63140 SKT. Ability as measured by the Mechanical AI apparently has the same relationship with SKT scores for both OJT and TTS graduates in this specialty.

The same results obtain for test specialty 64740, Materiel Facilities Specialist, except the difference in mean scores as evidenced by the "OJT" coefficient is slightly larger. The display would be similar to Figure 1.

Test specialty 70240, Administrative Specialist, is interesting in comparison because not only is the mean SKT score higher for "OJT" graduates, but also ability may have a differential impact on OJT and TTS graduates. Figure 2 illustrates these results.

Figure 1
FUELS SPECIALIST

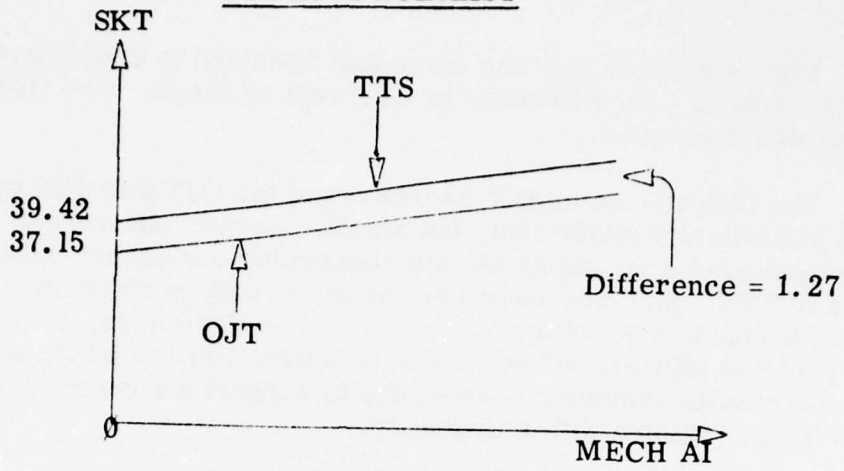


Figure 2
ADMINISTRATIVE SPECIALIST

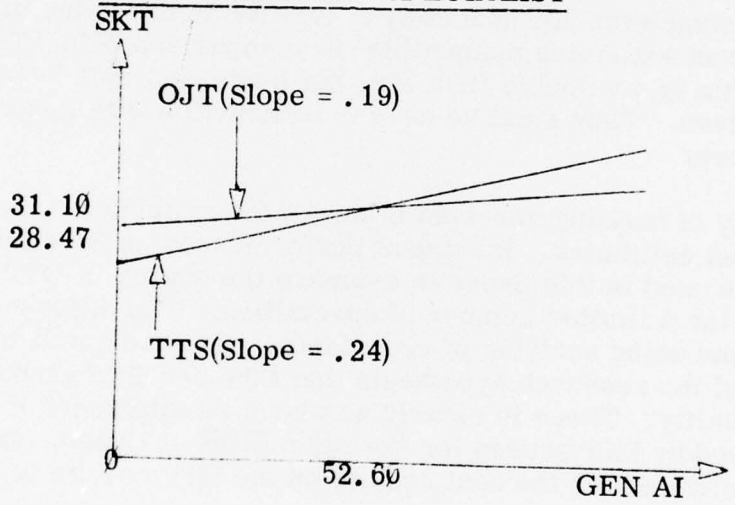


Figure 3
ALL OTHER SPECIALTIES ANALYZED

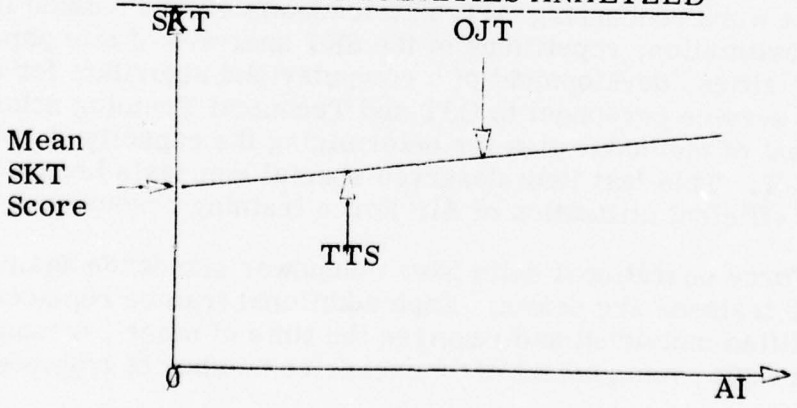


Figure 3 shows how the other test specialties would have appeared on the graphs - no difference in intercept or slope. The OJT and TTS lines are coincident.

The differences in SKT scores found for OJT and TTS graduates are statistically significant, but small. Larger differences might have justified a strong statement concerning the quality of OJT relative to TTS. SKT scores at best measure only performance potential without considering other components of performance, so small differences in SKT scores by themselves have limited implications. These results provide no evidence to support a contention that OJT and TTS graduates differ in quality.

V. CONCLUSIONS

Work completed and underway at AFHRL is providing the Air Force with OJT cost estimates compatible for comparison with TTS cost estimates. This is a valuable first step but these cost data have limited use by themselves. They must be used in conjunction with measures of other criteria.

Quality of training received is one of the criteria not incorporated into the cost estimates. Individual performances on Specialty Knowledge Tests were used in this paper to compare the quality of OJT and TTS graduates for a limited number of specialties. The differences in SKT scores found using analysis of covariance, were too small to allow rejection of the research hypothesis that OJT and TTS graduates are of equal quality. There is clearly no alarming difference in quality as measured by SKT scores for the specialties analyzed. One may not generalize either the cost results on the SKT results to other specialties.

Future work concerning OJT/TTS tradeoffs should include improved OJT cost estimation, repetitions of the SKT analyses of this paper for other specialties, development of a computerized algorithm for assigning non-prior service personnel to OJT and Technical Training Schools, and development of methodologies for determining the capacity of units to conduct OJT. This last item deserves special emphasis because it is the key to efficient utilization of Air Force training resources.

Air Force operational units have manpower standards against which OJT trainees are drawn. Each additional trainee replaces a more qualified individual and requires the time of other personnel to conduct OJT. Assignment of an excessive number of trainees

would result in unacceptable degradation of unit productivity. Even if OJT were less "costly" and produced equally well "qualified" graduates, there is a limit beyond which either TTS would have to be used or else manning standards would have to be changed.

Training costs form a large part of the DoD budget. Since the near future holds only budget cuts, one may anticipate increased interest in establishing procedures and data bases for selecting optimal OJT/TTS mixes. AFHRL research in this area provides a sound basis for continuing development of these methodologies.

REFERENCES

Dunham, Alan D. - The Estimated Cost of On-the-Job Training to the 3-Skill Level in the Communications Center Operations Specialty AFHRL TR 72-56, Lackland AFB, Texas: Personnel Research Division, Air Force Human Resources Laboratory, June, 1972.

Bureau of Naval Personnel - Manpower Allocation and Productivity Measurement Models, AD 701 300, 1969.

Johnston, J. - Econometric Methods, 2nd Edition, New York McGraw-Hill, 1972.

Lecznar, William B. - The Road to Work: Technical School Training or Directed Duty Assignment? AFHRL-TR-72-29, Lackland AFB, Texas: Personnel Research Division, Air Force Human Resources Laboratory, April, 1972.

Meade, D.F., & Christol, R.E. - Development of a Constant Standard Weight Equation for Evaluating Job Difficulty, AFHRL-TR-70-444, AD 720-255. Lackland AFB, Texas: Personnel Division, Air Force Human Resources Laboratory, November 1970.

Wonnacott, R.J. & Wonnacott, T.H. - Econometrics, New York: John Wiley & Sons, Inc., 1970.

APPENDIX A

OJT Cost Survey as sent to USAF supervisors with trainees upgrading to the 3(semiskilled)-level in the first five specialties listed in Table 8. The surveys were administered during the third quarter of 1973.

AIR FORCE

HUMAN RESOURCES LABORATORY

OJT COST SURVEY

AFPT 80-5X6X-109

This survey is part of AF Contract #F41609-72-C-0048, for which Personnel Research Division of the AF Human Resources Lab is the contract monitor.

AIR FORCE SYSTEMS COMMAND

BROOKS AIR FORCE BASE, TEXAS

DEPARTMENT OF THE AIR FORCE
AFHRL PERSONNEL RESEARCH DIVISION (AFSC)
LACKLAND AIR FORCE BASE, TEXAS 78236

REPLY TO ATTN OF: PESE (Capt Dunham, 4106)

SUBJECT: OJT Cost Survey

MAY 31 1973

TO: OJT Supervisors

1. The purpose of the attached survey(s) is to collect data concerning On-the-Job Training to the 3-skill level. This survey data, along with information from other sources, will be used in decisions concerning OJT and Technical Training School.

2. Answering the survey questions with some thought and effort will aid Air Force decision makers in the management of your AFSC.

3. Permission to conduct this survey was granted by Hq USAF/DPXOS, reference Air Force Personnel Test (AFPT) Number 80-5X6X-109.

FOR THE COMMANDER

OSCAR A. BERTHOLD, Colonel, USAF
Chief, Personnel Research Division

1 Atch
Survey

INSTRUCTIONS TO OJT SUPERVISORS

The accompanying survey is part of a research effort directed toward evaluating the costs and benefits of "On the Job Training." Your cooperation in completing the survey is requested. While it will probably take less than an hour of your time, the information you provide will be very valuable to the research and will help to improve Air Force policies concerning OJT and Technical Training School.

If you do not quite understand a question, give the best answer you can and feel free to write in an explanatory comment next to the question or on the back of the form. If you are completely uncertain about what a question means, enter a "?". If a question, for some reason, does not apply to your unit, enter "N.A."

The survey is divided into two parts: A and B. Part A asks you to try to make the best estimates you can about your average experience.

Part B asks you to keep a record of activities, each day for a week. It is important that you do this daily, so that what was actually done is fresh in everyone's mind. If you also feel that the week you reported on is not representative of your normal operations, so indicate by writing in an appropriate comment; and if you can, indicate what the average value ought to be in your judgment.

If you have any questions, contact Capt Dunham, Autovon 473-4106.

Q Number [| | | | |]

Upgrade Training in AFSC [| | | | | | |]

SPECIAL INSTRUCTIONS

1. The trainee's supervisor should complete this survey. Approximately one half (1/2) hour will be required to complete Part A, and five minutes per day for a week will be necessary for Part B.

2. When answering the questions, be sure to have a Job Proficiency Guide (STS), and the Consolidated Training Record AF-623 for each person undergoing training, handy to refer to.

3. The person who fills out this survey is encouraged to ask for the help of others, such as the OJT Monitor or an instructor when uncertain about the answer to a question.

4. Part A which should be completed immediately, is to be returned together with Part B within 8 days. Do not start Part B before completing Part A.

5. If there is difficulty in deciding what information is being asked for in any question, contact Capt Dunham, Autovon 473-4106.

BACKGROUND INFORMATION

NAME [| | | | | | | | | | | | |] [ ] [ ]

Last First Initial Middle Initial

GRADE If Air Force NCO enter "4" for E4, "5" for E5, etc. [ ]
If Air Force Officer enter "0."
If Civilian enter last digit of GS GRADE, e.g., "1" for GS.11.

SOCIAL SECURITY NUMBER [| | |] [| |] [| | | |]

PAS CODE [| | |]

PART A

1. How many trainees do you have upgrading to the 3 and 5 level in your section?

☐ 3 level trainees ☐ 5 level trainees

2. When a man (or woman) first reports directly from Basic Military Training, it may take some time before he actually begins training and work, even though his "date of entry" to training may be the same as his reporting date. This delay may be due to personnel processing, the need to wait for security clearance, or some other cause. Approximately how many days does it take before the newly arrived "helper" actually begins OJT?

☐☐ work days

3. There is also delay in entering training associated with the arrival of a 3 level from Technical School. In addition to personnel processing, familiarization with procedures specific to your situation may be necessary before he/she actually begins 5 level training. On the average, this delay is:

☐☐ work days

4. On the average, how many weeks elapse between achievement of 3 level proficiency and actual award of the 3 skill level AFSC?

☐☐ weeks

5. What week of training is your most average 1 level (helper) in?

☐☐ week

6. What % of the 3 level proficiency training do you estimate he has completed?

☐☐ %

7. When he arrived what % of the duties of a 3 level could he complete?

☐☐ %

8. If you stopped doing OJT training would you be able to reduce the number of NCO's in your work area without significantly reducing effectiveness? (Insert a "1" for Yes, or a "0" for No).

☐

9. During the training period for 3 level OJT, the instructor (trainer) must spend some time keeping training records up to date. On the average over the whole training period, how many hours (or fractions of hours) per week does the instructor (trainer) spend in record keeping for one trainee?

☐☐ Hrs.

GO ON TO NEXT PAGE

Part A (cont'd)

10. The newly arrived Tech School-trained 3 level is not as productive at first as the OJT-trained 3 level is, although he may soon close the gap.

a. In your estimate, what percentage of the workload of an OJT-trained 3 level can the Tech School graduate handle immediately after his arrival? ☐☐ Percent

b. How many weeks does it take before the Tech School-trained 3 level works with as little supervision as an OJT-trained 3 level? ☐☐ Weeks

c. After both types of 3 levels are awarded their 5 level, on the average do you consider either to have superior performance? (Insert a "1" for Yes, or a "0" for No.) ☐

d. If your answer was "yes," which type of 3 level do you consider to have better performance? (Insert a "1" for OJT, or a "0" for Tech School.) ☐

11. If extra (remedial) training is conducted in your office for trainees who fail the End of Course Exam (Apprentice Knowledge Test), answer the following questions:

a. In your experience, what percent of the 3 level trainees fail the End of Course Test the first time they take it? ☐☐ Percent

b. On the average, how many weeks of additional training are given to airmen who fail the End of Course Exam before they take the test again? ☐☐ Weeks

c. How many hours per week, during the normal work week, does the trainee spend engaged in this remedial training? ☐☐ Hrs.

d. How many hours per week, during the normal work week, does the instructor spend engaged in this remedial training? ☐☐ Hrs.

12. If you stopped doing OJT training and had no replacements for the trainees could your section continue to perform its mission without significantly reducing effectiveness? (Insert a "1" for Yes, or a "0" for No.) ☐

GO ON TO NEXT PAGE

Q Number [ ][ ][ ][ ][ ]

Part A (cont'd)

13. Based on your past experience, and, if you feel you need help, the experience of other qualified personnel in your section, list the average number of productive and non-productive hours of work for the trainee upgrading to the 3 level for each week between start of training and award of skill level. For instance, in the fourth week of training your trainee spent approximately 30 hours receiving instruction and reading and 10 hours doing productive work. Your second entry would look like this:

4	10	30

Note that the hours for each week must sum to 40, and you must have an entry in every week. If, on the average, trainees complete training between the 12th and 16th week, then the entry for the sixteenth week should show a "40" under productive and a "0" under instruction.

Weeks of Training (to the 3-level)	Trainee Productive Hrs Per Week	Instruction & Reading Hrs Per Week
1	[ ][ ]	[ ][ ]
4	[ ][ ]	[ ][ ]
8	[ ][ ]	[ ][ ]
12	[ ][ ]	[ ][ ]
16	[ ][ ]	[ ][ ]
20	[ ][ ]	[ ][ ]

14. What is the total number of personnel in your section (officer, enlisted, and civilian)? [ ][ ] no. of pers.

15. In addition to the trainees you now have responsibility for, how many more 1 level trainees could your section train right now without significantly reducing the effectiveness of section operations? (ignoring the limit on authorized number of personnel) [ ] 1 level trainees

16. If you had to lose a qualified 5 level for each new 1 level trainee (helper), how many more 1 level trainees could your unit train right now without significantly reducing the effectiveness of section operations? [ ] 1 level trainees

GO ON TO NEXT PAGE

PART B

FOR ONE WEEK PLEASE KEEP A RECORD AT THE END OF EACH DAY OF THE AMOUNT OF TIME SPENT IN EACH CATEGORY.

17. How many DDA airmen do you currently have enrolled in OJT to the 3 level? ☐ Airmen

18. Record daily the total hours your 1 level trainees spend on reading and receiving instruction each day. (Be sure and ask your trainees for their assistance in completing this question).

19. Record daily the total hours your 1 level trainees spend in activities contributing to office productivity.

20. Record daily the total hours of instruction provided by each grade of instructor.

	Mon	Tues	Wed	Thur	Fri
	☐☐	☐☐	☐☐	☐☐	☐☐
	☐☐	☐☐	☐☐	☐☐	☐☐
Q Number					☐☐☐☐☐
E3	☐☐	☐☐	☐☐	☐☐	☐☐
E4	☐☐	☐☐	☐☐	☐☐	☐☐
E5	☐☐	☐☐	☐☐	☐☐	☐☐
E6	☐☐	☐☐	☐☐	☐☐	☐☐
E7	☐☐	☐☐	☐☐	☐☐	☐☐
E8	☐☐	☐☐	☐☐	☐☐	☐☐
E9	☐☐	☐☐	☐☐	☐☐	☐☐
GS-5	☐☐	☐☐	☐☐	☐☐	☐☐
GS-6	☐☐	☐☐	☐☐	☐☐	☐☐
GS-7	☐☐	☐☐	☐☐	☐☐	☐☐
GS-8	☐☐	☐☐	☐☐	☐☐	☐☐
GS-9	☐☐	☐☐	☐☐	☐☐	☐☐
Officer	☐☐	☐☐	☐☐	☐☐	☐☐

AFTER YOU HAVE COMPLETED THE ENTRIES FOR FIVE DAYS RETURN THE SURVEY TO YOUR BASE CBPO

SYMPOSIUM

TRANSLATION OF TRAINING RESEARCH INTO TRAINING ACTION - A MISSING LINK

Introductory Remarks

G. Douglas Mayo
Naval Technical Training Command

I accepted the chairman's job in this symposium with the understanding that I would not have to be the kind of chairman who is neutral or impartial, but instead that I could expose my biases on the Translation of Training Research into Training Action, the same as anyone else. So, in introducing the topic it should not be surprising if I express some of the thoughts that I may have been repressing during the two decades that I have been involved in training research and in translating it into training action.

During this period, we have won a few. We are all familiar with instances in which the results of training research and development have been implemented into the on-going training operation and have contributed materially to it. Two examples that readily come to mind in Naval Technical Training (the area of operation with which I am most familiar) are programmed instruction during the decade of the 60's and the implementation of computer based instruction which is underway on a substantial scale at the present time.

But for every R&D project that has had an impact upon training there are numerous ones that have not. Now it can be argued that this is the nature of the beast, that risk taking is an inherent and necessary part of the R&D process. No doubt this is true, but I submit that the normal risks associated with R&D do not even begin to account for the unused, and in some instances unusable, volume of training research.

Much of the systematic civilian work in the area of translating educational research into educational change has been accomplished under the rubric of "linear change models in education." Most of these models describe a linear sequence of functions that include: research (the function in which new knowledge is produced), development (in which a product or procedure based on the new knowledge is engineered and evaluated), diffusion (in which the generality and extent of applicability of the product or procedure is explored), and adoption (the function of implementing an appropriate form of the product or procedure in an on-going educational situation). It is generally conceded that in order for such a model to work, at least one of two conditions must exist; either the people involved in the linear sequence of functions (i.e. research, development,

diffusion, and adoption) must be intrinsically motivated toward the common goal of improving education by means of research, or they must be extrinsically motivated by accountability with respect to accomplishing the common goal. More often than not, a combination of both intrinsic and extrinsic factors are at work in varying degrees in the instances in which the linear change model functions properly.

Moving to a more specific level, and with special emphasis upon the military situation, there are a number of conditions that contribute to the "missing link" in translating training research into training action. For convenience they may be divided into three groups, namely, those associated with research personnel, those associated with training personnel, and the interface between research personnel and training personnel.

First, research personnel. Not infrequently research personnel conceive of a "good" study and then look for a training situation in which to do it, without seriously considering whether the findings of the study, if successful, could be implemented in that or any other school. In my experience, it is not as unusual as one would wish for a researcher to complete a training research project in which his hypotheses are sustained, and yet be unable to state in what way his finding can be used to improve training or training efficiency.

A simple solution to the problem would seem to be to state in advance for each applied research project what action should be taken in the event that any one of several possible outcomes of the study should result. Carried to its logical conclusion this should prevent the researcher's wondering what to do with his results, upon completion of his research. Sometimes the researcher does not wonder what to do with his results because he does not feel that it is his responsibility to point out practical implications of his research, that his responsibility is limited to properly conceiving, designing, conducting, and reporting the study.

What about training personnel then? Are they not eagerly awaiting any research results that will improve their training operation? Well, yes and no, but mostly no. They are busy with the day to day pressures of conducting training. Both having the research conducted in their school and implementing the results tend to disrupt the training operation and pile additional work on a staff that already feels that it is overloaded. Often training personnel feel that the R&D was not their project, that any credit that might accrue from it is likely to go to the researchers or to a higher echelon in the training organization, and it is just hard for them to get very enthusiastic about it. Besides, several months often elapse between the time the data were collected and the time the more or less unintelligible report arrives.

This leads us to our third set of conditions that tends to affect the production of usable research and the proper use of it, the interface between research personnel and training personnel. This may well be the area in which the greatest source of our trouble lies. Too often research personnel do not accompany the research report when it is submitted to training people, do not make a presentation to training people concerning it, and do not assist in its implementation.

Clearly, it takes much more initiative on the part of the training people to decide to make changes in their training operation on the basis of a report than it does to reach a decision to take action when the author of the report describes the study, clarifies points that are not understood, and makes logical recommendations in a face to face situation to the people who are responsible for efficient training in the course concerned. The probability of successful implementation is further enhanced if the researcher continues to work with cognizant training personnel throughout the implementation of the research results. Normally this does not require a great deal of time and in my view, at least, is time better spent than in utilizing this same time on another research study, with the probability that the results of neither of the two studies will be implemented.

We have long known that in implementing any concept, responsibility for its occurring must be fixed. I am therefore suggesting that each organization engaged in applied R&D should have a small group, reporting directly to the Technical Director or to the Commanding Officer, whose primary responsibility is two-fold, first, to look externally to implementation of R&D products and, second, to look internally to the design and production of products that are capable of being implemented. Thus, on the one hand this group would ensure that R&D products were properly presented to appropriate training personnel, normally by the research personnel who did the work, and that assistance in implementing the products was continued as long as appropriate. On the other hand, prior to approval of an applied research project the group would assess the probable saleability of the product that the R&D project was likely to produce and so advise the Technical Director or Commanding Officer. If the purpose of the project was to produce a usable product and the proposed project was judged not to be capable of producing such a product, it would be referred back to the originator for revision.

At this point I will give the other members of the panel a chance to express their thoughts, repressed or otherwise. I hope that they are going to tell us that I am the only one who has observed the problems I have related, that everything runs smoothly in the training world with which they are familiar. But we will have to wait and see if that is in fact the case. The plan for the symposium is to have two prepared papers, one by a trainer (Mr. Walter McDowell) and one by a researcher (Dr. Norman Kerr). The papers will then be discussed by Lieutenant Colonel Donald Mead and Dr. William Moonan.

In addition, if I have said anything in my introductory remarks that is worth disagreeing with or otherwise commenting on, I hope that Colonel Mead and Dr. Moonan will include this in their discussion also. Following the papers by the two discussants, we will afford an opportunity to the four members of our Reaction Panel to make a few impromptu remarks. Upon completion of this structured portion, we will be pleased to open up the symposium for comments or questions from the floor. Our first speaker is Mr. Walter E. McDowell from the Army Training and Doctrine Command at Fort Monroe. He will be speaking on the basic topic of Translation of Training Research into Training Action, from the Viewpoint of a Training Manager.

SYMPOSIUM

"TRANSLATION OF TRAINING RESEARCH INTO TRAINING ACTION - A MISSING LINK?"

From the Viewpoint of a Training Manager

Mr. Walter E. McDowell
Supervisory Educational Specialist
Headquarters US Army Training and Doctrine Command
Fort Monroe, Virginia

When given the requirement to propose Human Factors-Behavioral Science research for the coming year, we are faced with a soul-searching situation in which we must ask ourselves two questions: one, are we really getting the taxpayers' dollar benefit from the work that has transpired in the past; and, two, will the research we are asking for be likely to yield productive results in the future. Another question which also must be scrutinized when it surfaces is, if we aren't getting a maximum return on our investment in research, how can we, as trainers and managers, translate the results of future research into better returns on our investment dollars? First of all, it is not purely a translation problem. The process of bridging the gap between the conduct of research and the implementation of research, by the Army trainer, is a continuing one. It begins while the problem is being defined to both the researchers and the trainers; it continues during the conduct of the research; it continues while the final report is being drafted, and it still continues long after the final report has been published.

Let's begin at the beginning. In a very real sense, utilization begins at the time the problem to be researched is presented. The training problem must be carefully defined and documented to insure that both the researcher and the trainer are speaking the same language and that the researcher fully understands the problem that the trainer believes he has. After the trainer submits his request for research, the researcher must then submit a statement to the trainer on what he views the problem to be and how he intends to seek out the solution. A common understanding and agreement of the problem to be researched is essential if a usable end-product is to be the result. The research approach may have to be modified in the light of continuing support requirements. Sometimes it may have to be recast if the research proposed is not understood by the trainer to be appropriately responsive to the problem as he perceives it.

Constant interaction between researcher and trainer takes place during the conduct of the work. Communication is achieved informally on a continuing basis through the use of interim progress reports, periodic reviews, briefings, and discussions. The trainer requesting the research must be kept constantly aware of the work taking place, and the researcher must be kept abreast of significant military changes within the parameters of training problems. Hopefully, the interaction and open communication will keep the evolving solution directly aligned with the trainer's problem.

As relevant data becomes available, that is, data which bears upon the problem, it is provided to the trainer. Such communication involves both those management functions which can implement change and those responsible for actually carrying out the decision to make the change. Management, simply decreeing that change be made, will not, alone, get the job done; the "not-invented here" syndrome must be overcome. The "doer"--the guy at the working level--must be convinced that the change is good and that it will be of benefit, not only to himself, but to the prospective students and to the system, otherwise it will be doomed to failure.

When the report is submitted, it is reviewed and staffed with appropriate elements of the organization. Each staff element examines and makes recommendations as to the suitability of the report as a research document. Some results which come from research are not usable in the formal report form because they may have been designed for background or for reference use, or perhaps to provide a basis for a review of training policies or programs. In all too many cases, the findings must first be subjected to a special translation and adaptation in order to become operationally useful. This translation is accomplished best by a joint effort between researcher and trainer and must be done with extreme care to insure that the research results are not lost, obscured, misused, or inoperable and at the same time are understandable and usable at the lowest level of organization-the units in the field-and by individuals who maybe limited in their intellectual ability.

Frequently it is essential to involve the research team in active participation in this translation effort. Team members provide assistance in the preparation of action directives, pamphlets, regulations and in adapting the findings to specfic operational contexts. Such is the case, at the present time, with the Fuctional Literacy Training (FLIT) effort at Fort Ord, California, and the ATC PERFORM project, a performance-oriented training program, being adapted at several training centers and the WAC School at Fort McClellan, Alabama.

Research by-products, such as experimental course units or programs of instruction, work books, prototype training aids and devices, programed text booklets and supplemental training literature, are similarly reviewed. It is considered that the incidence of these by-products being incorporated into on-going Army training efforts occurs with great and increasing frequency.

Utilization of research many times occurs while the research effort is in progress and prior to completion of the project. When the school or training center that has research sponsorship is impressed with, and convinced of the value of the work, on-the-spot implementation may occur through incorporation of those complete segments of the research results into the on-going training programs.

Results of research completed some years in the past are often found to be relevant to today's problems. How to tap the vast storehouse of past research to preclude re-inventing the wheel, is an area that deserves our immediate attention. Computer and other retrieval technology has been with us for some time. We need an efficient means of identifying past efforts as they may apply to todays problems.

Essential to utilization is the dissemination of research findings among potential users. The Training and Doctrine Command at Fort Monroe, Virginia, employs several means to alert elements of the command to research work, both in-progress and already completed. Bulletins to headquarters elements, staff briefings by teams representing the Human Research Units, the researchers themselves, the schools and training centers, and newsletters to headquarters staffs, centers and schools--all serve to publicize the results of research. Written quarterly progress reports are submitted and final research reports are appropriately distributed, including their entry into the Defense Documentation Center (DDC) and the Department of Commerce Clearinghouse for Federal

Scientific and Technical Information (CFSTI). All important are the informal and continuing contacts between researchers and potential users.

A report by Dr. William A. McClelland, at the Conference on Social Research and Military Management at the University of Chicago, June 1967, gives an insight into possible characteristics of unsuccessful and successful research implementation: UNSUCCESSFUL RESEARCH EFFORTS:

(1) Poor communication. Neither briefings nor reports effectively communicated the validity and operational value of the research.

(2) Lack of timeliness. The product of the research effort did not meet a valid, contemporary requirement. It was available too late or too early, or it was too tangential in nature.

(3) Degree of change. Too many changes in operating procedures were required. For example, training was shortened (or lengthened) too much, or the existing Army structure was incompatible with the indicated change. Existing or traditional practice may have been too strongly threatened.

(4) Lack of strong command support. Not enough people at high enough echelons wanted to change.

(5) Costs. Funds and personnel required had not been programed and could not be obtained.

(6) Lack of engineering capability. The training experts required to translate the research findings into more operationally usable form did not exist or were not available.

(7) Policy problem. There was a lack of doctrine under which to fit a new or improved training or operational capability.

(8) Insufficient "salesmanship". Project people did not devote enough effort to "selling" the product. At one time, for example, we believed this was not the job of the research agency.

Possible reasons for <u>successful implementation</u> are largely the obverse of this list. SUCCESSFUL RESEARCH EFFORTS:

(1) Timeliness. A recognized instructional gap was filled. The work was obviously relevant to a planned or on-going revision in Army practice.

(2) Command interest. There was a strong operational command interest, including that of a subordinate command. Interest existed at both management and working levels.

(3) Engineered product. The end-product was concrete. It was a material, plug-in item, specifically engineered for a given situation requiring little

additional Army effort to adapt it to the operational setting and requiring no doctrinal changes.

(4) Earlier acceptance by others. Some other service or civilian institution had accepted and successfully used the product or a very similar one.

(5) Personal interest. An individual Army officer or group of officers or key civilians associated with the work were convinced of its worth and were willing to serve as forceful proponents.

SYMPOSIUM

"TRANSLATION OF TRAINING RESEARCH INTO TRAINING ACTION - A MISSING LINK?"

From the Viewpoint of a Training Researcher

Dr. Norman J. Kerr
Director of Research
Naval Technical Training Command
Naval Air Station Memphis
Millington, Tennessee

I can never help but note that, each time I analyze the area of training research with respect to training management, the former comes out at the lower end of the continuum of power. There are reasons for this and not necessarily objective ones. And I tend to heap a prodigious quantity of the blame for this squarely upon the shoulders of the training management community. Too often, well-conducted programs of research have the result that the researcher becomes impaled upon the sharpened fence posts of ineptitude, inaction, and sheer ignorance. It also seems, at times, that well-intentioned research easily falls victim to a well-organized "dirty tricks" group of a well-endowed power organization.

Those of us in training research, while not necessarily comfortable with this situation have, none-the-less, grown accustomed to such conditions and are quick to sense disparities at all levels of interaction in the courses of the relationship between training management and training research.

It is precisely this disparity, it seems to me, that is the central focus for the consideration of this symposium. Why is it that training research and training researchers--and there is an overwhelming amount of evidence that this also might be said for the behavioral science field in toto--are not provided with the proper degree of status? In a society in which one is so willing to accept certain "truths", however unreasonable they may be, that flow from others who should know (National politicians are a good example), why is it that there is any difficulty to the acceptance of not only the word but the irrefutable data of the training researcher?

Perhaps those of you in training management think I do you some injustice. I cannot agree. The training manager, in my perception of him, always seeks answers yesterday to tomorrow's problems which he is causing today. He traditionally accepts a truism such as "individualization is 'good'" or "training time must be reduced by 10%" and, with unquestioning blindness, blunders forward with a series of compliance actions. Unquestionably, he must instinctively know that any adjustments to the system will result in some varying degree of turmoil. But it must be that he hopes that the turmoil will be temporary rather than permanent--and he does not willingly accept the fact that such actions may be not only intemperate or maltimed but are nothing short of downright wrong. It might be that he is insecure in his role and is often driven by power beyond his capabilities to control (much less to comprehend), be it national policies for the recruitment of personnel or service policies which dictate economic controls of severe stringency. He accepts the truism or the dictum and hastens to react. Perhaps he is intimidated by the world of reality as he perceives it and he feels that he has to remain flexible enough to get his job done in the face of such factors. He is, in actuality, a great compromiser; he plugs a hole here, redirects a pipeline flow there, introduces an innovation now and then, and follows the rules--those almighty rules--of action, and keeps his areas of interest both proper and orderly.

While the first one to say that job analysis must be the first step in the construction of an effective curriculum, the training manager is also known to be capable of putting together a curriculum on a complex

weapons system (for example) over night.

The typical training manager will fight the instructor billet battle seemingly to the wire by a continual wailing over inadequate numbers of instructors; yet he is very willing to accept new curriculum administration responsibilities for his overworked school staffs. Those who question this should ask themselves how many or in what proportions were new instructor billets provided when additional curriculum burdens such as human relations, defensive driving, and remedial reading programs were directed for inclusion into existing training programs. Yet, these good men seem obsessed with management and its precise controls, and often decline well-intentioned advice and assistance. That is putting it mildly, for it is only with reluctance and suspicion that many training managers will accept a rationale or the experience of others who are not only directly in but senior within his line organizational structure.

It might be that the training manager has the utmost confidence in himself--not necessarily in his abilities to improve the training, but most assuredly in his moxie to be able to handle the many and varied responses he will derive from both judgmentally sound and highly questionable decisions which he has made in the name of training management. After all, is this not what training management is all about?

Have I left out planning--or the lack thereof--by the training manager? I have not meant to. As any manager knows, planning is an essential part of his daily routine. My Lord, does he plan! He has training plans, personnel pipeline plans, SER plans, CIVSUB plans. . . Heaven knows how many other schemata he considers in the laying out of training requirements and in the execution of training activities. But ask the training management community for research plans and stand by for a high fog count in the responses. Better yet, ask for definitions of training research and you're likely to get responses in the order of "something not needed", "something received after the fact", and "something that is needed to verify my actions." The training manager, often willing to spend hours grumbling, perhaps, but unquestionably counting the number of sessions in a score of curricula to develop data to satisfy an obsolete report on a hands-on vs theory ratio, is generally not prone to review the research conducted on highly pertinent areas of his responsibility. Yet that research which might well enable him to better conduct the business of training may be in evidence all about him, gathering the patina of GSA dust.

Early in preparation for this symposium I decided that my assigned topic "From the Viewpoint of a Training Researcher" gave me wide latitude in pinpointing--at least to the satisfaction of my own perceptions--where the problem exists in the translation of training research into training actions. I think that I have now made that point several times over and I supposed that, since I probably have a goodly percentage of my listeners in a state of spirited agitation, I should smile broadly and back slowly out of the nearest exit. I promise that I will do just that very soon, but first let me say that I tried to locate strong support for my contention from the literature.

To my dismay I found that many of the articles on the subject in research and development professional journals are really self-deprecating. We in research are really rather prone to take on the problems of the world; to blame ourselves for another's inability or unwillingness to comprehend. The chest beating is often, in reality, a hushed humble breast-thumping which marks the contrite, and there are many in research who continually try to find the missing link of translating training research into training action alone. It was amazing how sparse was the material in support of the rather parochial position I have presented. As a matter of fact, I found the strongest proponent of the view that management refuses to take seriously well done research in the management field. I certainly learned once again the lesson that one's own literature is the best source of criticism. Thus, if researchers want to feel badly about their role in the missing link bit, they merely have to turn to sources such as the American Education Research Association's Review of the Educational Research, Phi Delta Kappa's Kappan, and the like. Conversely, the management literature produces gems designed to make researchers feel good (if not full of self-pity).

Since I enjoy feeling good, I picked up a copy of the Industrial Research Institute's Research Management and found a convincing position I would commend to you at this time.

Stanley I. Berman, Vice President of Manufacturing and Abrasive Operations for the Norton Company, in the July 1973 issue of Research Management has an article, the title of which initially intimidated me (as a researcher) considerably. It is "Integrating the R&D Department into the Business Team." Mr. Berman, not at all abrasive (as it turned out) contends that management has continually not been able to conceptualize the need and value of research. He claims it is the fault of management that there is a breakdown in the production team. He states: "....management....traditionally has not known what to expect from researchers, and accordingly has given them little direction."

He goes on to say that management needs to tell researchers in detail what it requires, evaluate the researchers capabilities to accomplish the objectives and then ".....if feasible, require them to do it."

Mr. Berman notes that management has not supported research because it has traditionally not participated in R&D, and has not understood how it can support R&D objectives. He contends that there needs to be the concept of integrating research and development into total corporate allocations--the business plan, in effect. And, further, the R&D organization needs to be provided with a careful definition of the problem faced and the solutions required. Let me quote Berman again:

> "If we don't do this (providing the careful definition and the solutions required), if we continue to allow a situation to exist where we are constantly wondering how much support we are giving R&D, then we are openly admitting that we as managers have not succeeded in making research an integral part of the business system and we are risking total business failure."

So, you see, the solution to the problem is clearly in the hands of management.

Now, being in research I cannot help but come forward with not one but several approaches for solution.

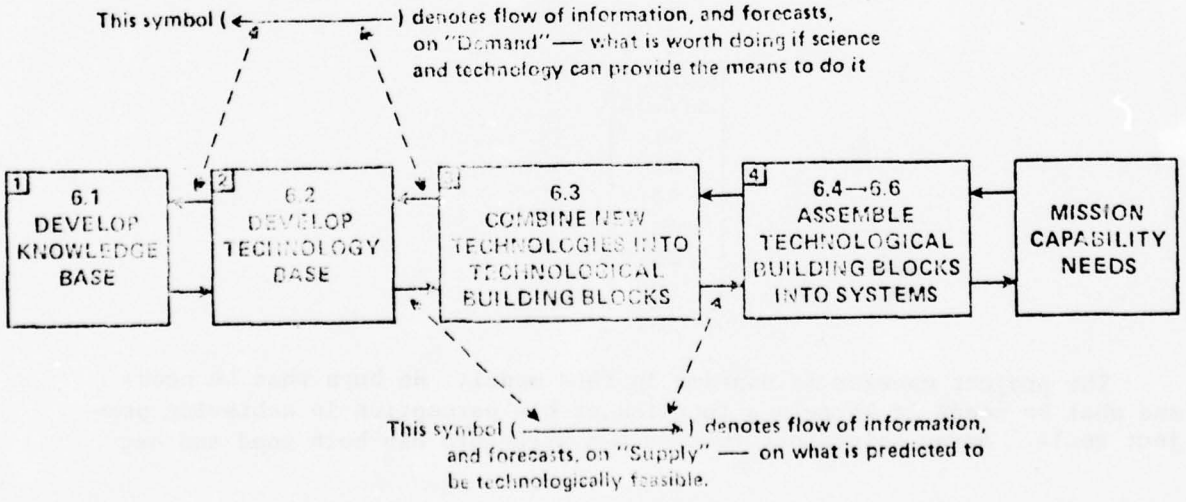
The problem is not easy to solve; there are many levels of competencies as well as many specialties in the research function just as, in production facility, one will find a variety of competencies and specialties within the work force. In the research function, for example, there are personnel who work well in fundamental research problems but who, perhaps, are impractical in the broad field of applied research. The converse is axiomatic. The Department of Defense RDT&E structure speaks to such specialties in the functional terms of:

6.1 Fundamental research
6.2 Exploratory development
6.3 Advanced development
6.4 Engineering development
6.5 Management and support, and
6.6 Operational Systems development,

recognizing the unique facets of research and development. It seems to me that management should recognize, if not the categories, then certainly the functions of the several types of research programs.

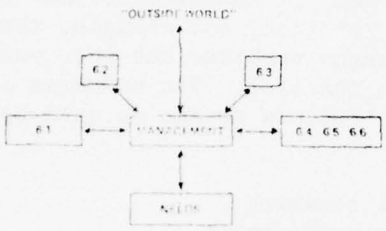
Glenn Bryan, in a published article entitled "The Role of Basic Research in the Total R&D Process", in the January 1973 issue of Research Management, conceptualizes a series of models which address the enigma of placement of the various types of research into an organization.

Among the models are several organizational ones with which we are familiar. First, he depicts a "Linear Production Model" which, for my purposes today, adapts the DOD research functions into a production scheme.



You can see, the block to the right draws upon and feeds responses to the several research functions to the left. This situation is somewhat idealized. It appears in the straightforward approach that flow to and from the blocks is continuous, unimpeded, and simple. It would be nice were that so.

Another model of Bryan's is the "Departmental Model."



This model has decentralized the research function and superior authority which Bryan calls "Management" in the model takes total responsibility for coordinating the several research functions and moderating communication among them. Bryan states that such a system as depicted in the model requires superior technical competence at the management level to function effectively. I might also add that it will take superior managerial competence to keep such a competitive system functioning toward the overall goals of management.

Another model is the "Project Model" which has had much usage throughout the Defense establishment.



The project manager is supreme in this model. He buys what he needs and what he needs is largely a function of his perception in achieving project goals. Bryan points out that such a structure has both good and bad

merits. Inasmuch as the highly focused research efforts, that are in line with project goals, are often well supported since they are usually an adjunct to a high-visibility, large project, that is good. The drawback, of course, is that there is little general support for research not directly in evidence of project goals. Research runs the constant risk of being a necessary evil in this model.

I won't show you Bryan's "Organic Model." There is little need to, for consider a tree, if you will, laden with lush fruits. Then consider its root structure, seemingly disorganized, ugly, and dirty. Then consider the roots as what they are, the source of the healthy fruit; the roots that are often misunderstood as to their function and their problems. That the roots can isolate the proper nutrients that are necessary for effective fruit production is without question. Also without question is the fact that indiscriminate root pruning will almost certainly have adverse effect upon the fruit.

Bryan's tree is more an analogy than a model. It is too simple to portray in any real sense of the world what the interaction between management and research is; it merely portrays what it should be.

And this, perhaps, is the reason that we live in a world of linears, departmental, and project models--variations, perhaps, of those described. A reasoned approach by management is necessary to bring about precisely that fine-grained relationship of mutual respect within organizations among its major functionaries. The cartoon showing the manager in his plush office speaking to a person in a laboratory coat, and bearing the caption: "You researchers--make some kind of a breakthrough!" has got to cease being funny.

It would seem to me that the understanding of the R&D function, the placement of that function within the dynamics of the organization, and the understanding by management that reasoned direction is required will bring the closure we all seek and eliminate the need to worry about the translation problems between training research and training actions.

SYMPOSIUM

TRANSLATION OF TRAINING RESEARCH INTO TRAINING ACTION--A MISSING LINK

DISCUSSION

Lt Col Donald F. Mead
Chief, Training Technology Division
Randolph Air Force Base, Texas

My charter this morning is to serve as a discussant of the somewhat diverse views expressed by Dr. Kerr, representing the training researchers, and Mr. McDowell, the training manager. Following Dr. Mayo's precedent I choose to exercise the prerogative of interjecting personal opinions when I feel they might stimulate thought or possible solutions to the problems encountered in the transition of research data into training programs. To assist you in weighing the value of my comments, I might point out that through military assignments and promotions I have progressed through the role of researcher into that of training research manager. Let me say that there is more than a "just noticeable difference" in the two environments.

After a careful review of the presentations by Dr. Kerr and Mr. McDowell it appears that the purposes of this seminar could best be served by focusing on their agreements rather than disagreements. Following this format, I will also discuss my personal reactions as they occur.

Of prime signficance to the military services is the agreement that a problem does exist in the transition of research products into training programs. This should trouble managers at all echelons since it indicates significant waste in personnel and financial resources. Impending reductions in DOD manning and funding will make this situation increasingly intolerable.

While our speakers agree that a transition problem does exist their perspective on the cause differs.

Dr. Kerr sees management as the culprit due to a skewed power structure and certain training managers who are inept, lazy, and ignorant. On occasion these managers ignore irrefutable research data, are suspicious of well intended advice or assistance, and fail to provide the researcher with direction in the form of a long range requirement plan. In view of Dr. Kerr's description, one has to question his seeking direction from such a group.

Dr. Mayo concurs that managers contribute to the problem and suggests that managers are generally not receptive to research due to job pressure, disruption of operations, additional impact on the staff, and failure to identify with the effort. He does, however, place part of the problem at the doorstep of the researcher. Specifically, he cites the scientist who conceives of a problem area, conducts a study, then seeks out a training situation for application. I think we are all familiar with this researcher whom we might liken to a free floating umbilical cord seeking a warm, receptive training course to plug into and thus nourish his creation. This fellow is only outdone by the research jockey who develops a research horse and rides it to death from iterative race to iterative race with little thought given to solving problems and meeting needs. The prime objective here appears to be another journal article or ticket to an upcoming conference or symposium.

Dr. Kerr suggests what I perceive to be the real key to moving research products from the laboratory to the classroom, and that is comprehensive training research plans and requirements. I believe the managers' suspicion, reluctance, and resistance may frequently be traced to research conceived in the isolation of a laboratory and subsequently offered to solve a training problem. Typically, this requires the original research product to be modified, adjusted, and expanded, with predictable results. It is not surprising that book shelves are loaded with research reports never applied. In many instances, I suspect it is because they were not designed in response to stated operational problems.

Along the same lines but somewhat deeper, I feel that major modifications must be made to gain coordinated management of the multiple DOD research agencies. Currently 6.1, 6.2, and 6.3 efforts are often initiated

without cross-validation or coordination with training managers. It is not unheard of to have a contractor, whose unsolicited proposal is funded with 6.1 monies, arrive at a military installation without announcement or coordination. Similarly, there are examples of duplicated efforts while topics needing exploratory research remain untouched. At this time the military services are open to criticism in this area. Personally I believe that no research, regardless of functional category 6.1, 6.2, or 6.3 should be undertaken unless it supports stated user requirements. Recently an Interservice Training Review Board has been established which could make significant contributions in this area.

While I agree with Dr. Kerr that a training research requirement document is an essential item, I don't believe that it is solely a management responsibility. Ideally this document would be developed in coordination with training researchers. Carried a step further, it could be a matrix consisting of a taxonomy of significant training research areas on one axis and research agencies on the other. When ongoing research efforts are plotted, those areas not being addressed become readily apparent.

At this point, I fully endorse the close coordination between researcher, manager, and trainer presented by Mr. McDowell. Several points of his presentation warrant reemphasis: A meeting between researcher and training manager should be held prior to initiation of the research; the statement of work should be coordinated and approved; proposed research products should be reviewed and approved as developed; and close interaction should be maintained as the project progresses. This coordinated spearhead will open a pathway through management's anti-research field of sharpened fence posts described by Dr. Kerr.

Another area of general agreement among the presentors is the problem of placement of R&D activities in the organizational structure. Dr. Mayo suggests the R&D group should report directly to the technical director or commanding officer. I concur in this suggestion. Placement in the upper echelons of the organizational structure is critical since it dictates the interface between the training manager and researcher. In addition, it eases access to top decision makers which is essential for support of research projects requiring major expenditures of resources or having significant impact on the training activity.

Because of several unique aspects a brief description of Air Training Command's R&D organizational structure might be of interest to this audience. It should be understood that the responsibility for personnel research in the Air Force has been assigned to the Human Resources Laboratory, Air Force Systems Command. Research requirements from each major command are forwarded to them for accomplishment.

Somewhat unusual is the fact that the HQ ATC R&D organization is located under the Deputy Chief of Staff for Plans rather than under the DCS for Technical Training or DCS for Operations, which involves flight and navigator training. This placement permits close interaction between future command programs and R&D requirements, permits more comprehensive scanning of state-of-the-art technology than might be possible under an agency responsible for specific training, and provides access to the commander.

Another unique aspect is the personnel make-up of the HQ R&D staff, which consists of a computer specialist, two pilot-flight simulator specialists, a navigator simulator specialist, an audio-visual specialist, and two behavioral scientists. Each specialist is charged to interact with academic, civilian, and military agencies developing products or conducting research in his area of responsibility. Information concerning new technologies with potential use in the command is coordinated with the appropriate DCS and/or forwarded to the appropriate training activity for coordination. Where interest is shown arrangements are made for visitations to the research agency or briefing of appropriate staff personnel by the researcher. Each specialist is also responsible for initiating Requests for Personnel Research in his area of interest to be conducted by the Human Resources Laboratory.

Training Research Applications Branches or TRABs have been established at each Technical Training Center, monitored by the HQ R&D group. The TRAB works for the school operations officer and has access to the school commander. Each TRAB has a civilian GS-11 educational specialist and two military behavioral scientists. These organizations perform quick reaction studies for training managers, conduct application studies with training researchers as directed by the HQ R&D group, and initiate research requests to be accomplished by HRL.

All TRAB projects and research requests are controlled by the HQ R&D group to prevent duplicated effort.

A Research Review Board made up of the technical directors from Technical Training, Flight Training, and Plans has been established to validate command research requests. In addition, they monitor the action taken on recommendations resulting from completed research studies. The Research Review Board responds directly to the Chief of Staff and the Vice Commander.

The final area of agreement mentioned by our presentors deals with the requirement for communications between researcher and training manager. If a research plan is developed and there is close coordination during the initiation, conduct, and reporting of research, communications problems are rarely experienced.

I have come to several conclusions regarding communications during the past three years of coordinating the interaction between researchers and high echelon decision makers. Some of my conclusions concur and others differ from the positions taken today by our speakers. My remarks reflect my personal opinions only and are specifically directed toward the interaction between the researcher and highest level decision makers.

Researchers are their own worst enemies in the transition of research products from the laboratory to the training environment. Specifically:

a. As a group they are particularly poor briefers. They generally arrive prepared to deliver a three-hour briefing to a person who has limited time for discussion. Their briefing format usually follows that of an APA journal article and is usually strongly statistically oriented. Although experts on their studies, they seldom address the critical aspects affecting the decision, i.e., the total resource impact of implementing their work. I recommend that prior to addressing a decision maker you work with his staff to develop a complete package and if possible have appropriate members of his staff present to field "in house" questions you can't answer. Design your presentation to answer the question, "What does this man need to know to make his decision?"

b. The publishing of selected bibliographies and reviews of literature may look great on the vita but it tarnishes the researcher's image among managers. I shudder when these documents reach the headquarters because they invariably precipitate two questions: "Is this what we are spending our money on?" and "What the hell am I supposed to do with this pile of junk?" This is particularly damaging when expectations are high concerning a new project and this type document arrives representing the initial work expended. My recommendation would be to hold the document and submit it at the conclusion of the project or incorporate the entries with your interim reports.

c. The time-honored format of the typical research report may be justified for publishing ease and communications between professional scientists, but it is ill-designed to convey meaningful information to training managers. Again I would encourage you to consider your audience and write to meet their needs. Minimize the discussion of your statistical design and consider presenting this information as an appendix to the report. I really recommend dropping the APA report format and going to executive summary type reporting.

d. And finally, when a fast reaction situation requires, stick your neck out. If a decision has to be made, don't suggest you can provide a solution in 18 months. Use your knowledge of research data, and come up with viable options. Validate as you go and modify the approach as required.

SYMPOSIUM

TRANSLATION OF TRAINING RESEARCH INTO TRAINING ACTION - A MISSING LINK

DISCUSSION AND COMMENTARY

EARL I. JONES
NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER
SAN DIEGO, CALIFORNIA 92152

Having reviewed with great interest the papers of Mayo, McDowell and Kerr I shall attempt to summarize their main points, comment upon them, and then as is a discussant's privilege, provide some comments of my own.

Dr. Mayo, with whom I have worked closely for the better part of two decades makes several major points. His first major point seems to me to be the rationale of this symposium. It is that a great amount of training research and development goes unused. He stipulates that failure to use R&D cannot be accounted for by the simple assumption that risk taking is a fundamental part of R&D nor by the fact that some of the unused training R&D is indeed unusable. He leaves out an important (but fortunately small) class, namely, training R&D that has been used but should not have been.

The next major point in Dr. Mayo's paper is his reference to "linear change models in education" with the stipulation that conditions of intrinsic or extrinsic motivation must be functioning for the model to work.

From here Dr. Mayo moves to the position that most of the conditions functioning to obviate use or translation of training R&D are associated with (1) research personnel, (2) training personnel, or (3) the interface or interaction between research personnel and training personnel. With respect to research personnel Dr. Mayo perceives the major sin to be conceiving studies in which researchers look for training situations in which to execute but fail to consider (in the planning stage) the utility or implementability of the R&D results. As a solution to this unwholesome state of affairs Dr. Mayo has the pragmatic notion that researchers should pre-state and pre-judge the training actions to be taken contingent upon the possible R&D outcomes and accept responsibility for such planning.

With reference to training personnel Dr. Mayo describes their work overload, their dislike of waiting forever for utilizable R&D results or recommendations, their commonly not identifying with the R&D task or project, and their feeling that whoever might get credit for research based action, they (the training personnel) will not.

Dr. Mayo sees the interface or interaction between researchers and trainers to be the most critical problem. Here he recognizes that the continuity from R&D to translation to change in training can be accomplished only by appropriate interaction in which researchers and their managers take the responsibility for the translation process and assist the training personnel in the action process to the point where appropriate implementation has been accomplished.

Mr. McDowell's presentation is a natural follow-on to Dr. Mayo's. He perceives the requirement for proposing research as a soul searching responsibility for use of the taxpayer's dollar involving the questions of benefits gained from past R&D and projected yield of future R&D. He perceives the trainers' and managers' roles as not limited to translation, but a process of bridging the gap between conduct of research and implementation. This process is continuous according to Mr. McDowell and requires constant interaction between researcher and trainer from problem identification, R&D planning, conduct of R&D through translation and implementation.

Mr. McDowell is not so prone as Dr. Mayo to perceive a vast graveyard of unused R&D. He qualifies his remarks in a way which implies that the Army may not suffer the "missing link" dilemma to the extent that the Navy does.

Where Dr. Mayo exhorts the need for trainer-researcher interaction, Mr. McDowell talks of a system in being. He refers to interaction that "takes place;" data which are "provided to the trainer;" reports which are "reviewed and staffed with appropriate elements of the organization;" and specific examples of frequent and essential involvement of the research team in "active participation." He perceives use of R&D by-products as increasing, and dissemination of R&D findings among potential users as essential.

In closing, Mr. McDowell excerpts from Dr. McClelland's 1967 report eight characteristics of unsuccessful research efforts and five characteristics of successful research efforts.

The final presentation by Dr. Norman Kerr is a shocker. In some communities it might even seem blasphemous. However, in the long course of the history of education, the history of training and the history of training R&D, Dr. Kerr's remarks spark a flame of truth--albeit a truth so mixed with exception, complication and dark shadow that its explication is no simple matter.

Dr. Kerr's paper states what the other papers only imply. The problem may not be a missing link but a faulty chain anchored insecurely and subject to vectors of power and stress that make the task of translating and applying R&D results in a positive manner unduly difficult or even impossible.

Dr. Kerr's major points are that training managers obviate rather than facilitate R&D, that they possess power whereas researchers do not, and that the power disparity at all levels of interaction between training managers and training researchers are the central focus for consideration in this symposium. Dr. Kerr carries this argument with conviction, verve, eloquence and numerous examples. Although he provides a scorching indictment of training managers, Dr. Kerr does not back slowly out of the nearest exit. He finds to his dismay that his review of research literature does not provide the confirmation he seeks but that management literature does. With support from research management literature and his own knowledge of the R&D community's management models, Dr. Kerr provides some positive proposals for rapprochement.

We now have summarized the major points made by the three speakers. Let us comment briefly upon them.

To my ear at least, the three papers we have heard are extremely useful accounts of the success or failure of military organizations to exploit R&D as a means of upgrading military training. As we might expect the papers strongly reflect both the nature and nurture of the problem and the histories of the presenters. I, too, will try to overcome that dilemma of all observers--how to get outside of my own skin while being securely bound within it.

First for Dr. Mayo. The explicit concurrence of both Mr. McDowell (by his reference to McClelland) and Dr. Kerr as a strong protagonist would be sufficient for me to accept the assumption on which this symposium is based; that is, that there is a significant amount of useful training R&D that does not get exploited by the military training community.

Time does not permit complete comment on "linear change models;" suffice it to say that they are useful devices which like most models have various constraints. The primary one is perhaps the need for intrinsic or extrinsic motivation. Here Mr. McDowell's cautionary observation that a decree by management (one type of extrinsic motivation) will not get the job done is fair warning and is echoed in different ways in both Dr. McClelland's list and in Dr. Kerr's admonitions. Extrinsic motivation is then sometimes necessary but seldom sufficient. However, given that one of the motivational conditions is necessary, the problem becomes one of controlling the direction of its outcome.

Within the context of this symposium, Dr. Mayo's point that translation of R&D into action depends upon research personnel, training personnel and their interaction is so strongly supported by McDowell, Kerr

and logic that it needs little commentary except to wonder why, when the problem can be cast into such simple terms, the solution to the problem has not long ago been routinely applied.

Mr. McDowell's paper is so convincingly positive that it would seem, for the Army at least, the solution is at hand.

Shifting now to Dr. Kerr provides a somewhat stark contrast. He ignites such a conflagration that we are heaped with ashes and wonder as in the "Perils of Pauline" if there can ever be a happy ending. By following Dr. Kerr further, the most obvious answer would appear to be a qualified "yes." Equitably distribute power, select a management model, integrate R&D into the team and, I will infer, we will be on our way to some kind of break through--not ordained and not funny.

To me, a bothersome aspect of Dr. Kerr's presentation is his broken field running. He, with some justification, puts the main burden on training managers. He finds researchers to be hushed, humble and contrite (hardly characteristic of his own presentation) and finds solace in management literature. His two major references are Berman and Bryan. Some comment is called for here. Both of these references deal not just with management but with research management. Berman's message, intended for the industrial world, is that the R&D component of a company should be part of the business team--to participate with and interact with management. For decades this has been true of such major companies as Bell Telephone, IBM, Union Carbide and countless others. Dr. Bryan's article was motivated by a three day R&D symposium concerned with the adequacy of DOD R&D structure and its implied lineal model for the management, planning, funding and execution of manpower and behavioral science R&D. Dr. Kerr uses these references to advantage in making his strongest positive point, the need for research and management organizations to become an integral force for product improvement and organization efficiency. Had Dr. Kerr's search of research literature been somewhat more extensive he would have found a good deal of support. The writings of Finn, Lumsdaine, Glaser, Rigney, Mackie, and even Earl Jones would not only be consoling but might dispel his notions of contrite.

If my interpretation of these papers has been correct, I heartily endorse not only their major aim but their explication of factors which obviate implementation of R&D into training action and their recommendations for positive steps for catalyzing, facilitating and improving the translation and implementation process. Yet I think they have not gone far enough. What seems to me to be missing are a set of models for training management, R&D management and a meta model, superordinate to training and R&D, which could guide both the training community and the R&D community in the spirit of integration, establish all the necessary interfaces at all levels of function in both training and training R&D and provide such mechanisms as a Training Research Advisory Board and laboratory schools--in short, a model which could facilitate the kinds of interactions all three of our speakers have proposed in their own unique styles.

TEST FEEDBACK AND TRAINING OBJECTIVES

BY

CARROLL H. GREENE, LTJG USCGR

UNITED STATES COAST GUARD TRAINING CENTER

CAPE MAY, N. J.

Note: The views expressed in this paper are those of the writer and do not necessarily reflect those of the USCG Training Center, Cape May, N.J.

INTRODUCTION

The public seems to be generally willing to accept a decrease in the manpower of the armed services. As a result of this acceptance the active draft system has been deleted as a method of acquiring servicemen. Whether this acceptance is warranted remains to be decided however, and in the meantime this decrease in manpower is something with which the military forces must deal. Why must we deal with it? Because, in the event that military or naval forces must be called to action on short notice, the American people will expect and demand the same readiness and response which has always been characteristic of the armed forces. An uneffective response at such a time could cause a breakdown in one or all of the systems which defend our National Security. It is apparent then that, in the face of major cutbacks in the quality and quantity of individuals joining the armed forces, we must continue to maintain an effective state of military readiness which is responsive to situation demands.

Many times we lose track of the overall importance of our jobs when we become involved in the tasks of everyday business. However such large problems as that mentioned above are not remedied by one or even several decisions or changes. The overall situation, in this case is affected by many small variables and solution to it will be dependent upon many small adjustments or changes which will in turn have an overall effect.

The situation with which this study deals is one of those small

areas in which changes can be made which will affect the ability and readiness of the men presently being trained by our armed forces.

To increase the abilities of our servicemen in their various specialties, as well as in the general knowledge of those things which pertain to the military forces, is an objective which will help accomplish the readiness which the public expects. This can be achieved even in the face of lower enlistment rates and manpower cutbacks. In increasing retention and usage of information given to trainees we can project that this ability will enable them to operate at an increased level of awareness, and efficency.

At the Coast Guard Training Center in Cape May New Jersey we often encourage instructors to give information to recruits in regard to performance both on the positive as well as negative side. This feedback is what most people base their evaluations of their abilities in any particular area on.

In an early experiment on feedback Hurlock (1925) emphasized encouragement or incentives and their effect on learning. He also encouraged the use of verbal praise or criticism and added power to the preference for verbal as opposed to other types of feedback. Plowman and Stroud (1942) showed that seeing their errors corrected on an examination allowed high school students to eliminate approximately 50% of their errors on a retest one week later. This result was probably due a great deal to the practice effect but does point to the value of feedback in enhancing ability. More recently Page (1958) reported

results which bear out Hurlock's earlier emphasis on praise and blame. He found that students who received comments on examination papers did better on the next examination then did students who received no comments.

Much of the previous research in this area has been conducted comparing feedback with no feedback situations. Curtis and Wood (1929) however, tested four methods of scoring examinations and found that those methods which provided for discussion of questions were the most effective. Stone (1955) found that the effect of feedback was correlated positively with the amount of information contained in the feedback. Students who received full explanation, as to why one answer was wrong and another was correct, did much better subsequently than did those who received less, or no, information. Sassenrath and Carverick (1965) compared four treatment groups for retention and transfer of learning and concluded that the type of feedback is not nearly as important as the fact that the group gets it. In the present experiment it was hypothesized that the introduction of weekly quizzes prior to the first progress test in recruit training would increase scores on the progress test and this increase would vary positively in accordance with the amount of feedback given on these quizzes.

METHOD

APPARATUS Quizzes consisting of twenty five questions each, based on the first and second week subject matter were constructed. Questions covered the same subject areas as those on the third week progress test but were changed to appear in a slightly different form.

All quizzes were given in well lighted classrooms under comfortable conditions.

SUBJECTS and DESIGN Five hundred and sixty eight Coast Guard recruits who entered the service between 15 April and 15 August 1973 served as subjects in this experiment. They were assigned to companies, approximately fifty men each, as they reported for duty. The varying numbers in each company caused the number of subjects in each treatment group to fluctuate. As each company of men began training they were assigned a treatment category, 1, 2, 3, or 4. This continued until four companies had gone through treatment 1, four companies had treatment 2, five companies had treatment 3, and four companies had treatment 4. The total number of subjects in each group was as follows.

Treatment #1 - 145

Treatment #2 - 134

Treatment #3 - 170

Treatment #4 - 119

568 = N

Subjects were therefore assigned as randomly as possible to treatment groups. Four instructors took turns administering the three treatments to help control for the effect of any particular instructor personality.

PROCEDURE Prior to this time no quizzes were given to recruits other than the third and sixth week progress tests which consist of 50 multiple choice questions on material covered in lecture, course material and text. Quizzes were introduced at the end of weeks one and two prior to the first progress test. Companies received their assigned feedback

condition over both quizzes and were afterwards tested using the standard "first progress test". The treatments for each group are specified below. Recruits were told that quiz grades would not be a part of their record and were for their information only.

TREATMENT #1

After the quiz each man kept his paper and the answers were reviewed. When no questions came up in regard to an item the instructor was told to explain the answer anyway. Recruits were encouraged to ask questions and discuss each one fully.

TREATMENT #2

The companies which took part in this treatment were told which choice was correct in each case, but were not given any further information or allowed to question it. Thus they received some feedback but not the amount available to treatment group one.

TREATMENT #3

Since we were introducing two independent variables in the feedback and the quiz itself, it was necessary to be able to say wether change was due to the feedback or merely the introduction of the quiz. It was for this reason that group three received only the quiz and no information as to how well they did on it. Instructors required that answer sheets be turned in upon completion and gave no information on correct answers. Subjects received no feedback other than how they felt they had done on the quiz.

TREATMENT #4

The control groups proceeded through training as other companies had until this time, with no quizzes until the progress test at the end of the third week of training.

RESULTS

A Kruskal - Wallis one way analysis of variance yielded an Hc value significant at above the .05 level. Group comparisons, again using the Kruskal-Wallis formula, were made which indicated significant differences between several of the specific groups.

INSERT FIGURE 1 HERE

These results imply that a definite difference existed between the group which received maximum feedback from the quizzes and both the control and the quiz - no feedback group. This difference bears out the hypothesis and was presumably due to the effects of feedback. The max feedback group did not however, differ significantly from the minimum feedback group (group II). In this respect the increase in feedback did not cause a significant difference from the lesser feedback level and fails to bear out the finer implications of the hypothesis. The trend was toward an increase in feedback predicting increase in performance, but in this case it did not approach significance.

DISCUSSION

The results of this experiment bear out the hypothesis in regard to feedback versus no feedback and seem to confirm a statement made previously (Sassenrath and Garverick, 1965) that the type or amount of feedback is not of extreme importance in these situations. The important thing is that there is some feedback. The data shows a trend, although not always at significant levels, that as the amount of feedback increases so does performance.

Instructors noted that throughout the experiment the groups which received max feedback showed a good deal of interest in other areas of training and asked questions which concerned more than just the quiz. One important factor in this case seems to be that of human interaction; the chance to speak with, or question, someone higher up in the chain of command. This interaction was seen to have an effect in terms of heightened interest in many areas of military life. These companies received the treatments twice over a two week period and the question crops up wether the difference in group one's performance over that of three and four was due to the educational aspects of feedback or the psychological aspect of interaction with someone of higher rank on a semi informal basis. Both factors probably affected the results but it would be interesting to see wether the feedback was valuable because of its information or because of the human interaction.

A certain type of practice effect probably occurred also in that a majority of these men indicate a fear or apprehension concerning test taking. Allowing them to take two quizzes prior to the progress test probably helped eliminate some of the anxiety normally surrounding this test and present in group four.

Due to the nature of the environment subjects in this experiment could not be assigned to treatment groups on a completely random basis. There is a fluctuation in recruit education level that occurs at various times throughout the year, and it is possible that this fluctuation may have had some effect on results. However companies were assigned a treatment group in sequence as they entered training and this fact should have ramdomized the effects of the seasonal fluctuations.

Whatever the reasons for the effects of feedback in various situations,

we know that it does enhance performance and increases retention of information to a recognizable extent. This realization should be of prime importance in developing any training program and should incorporate the use of human interaction in increasing a persons interest and knowledge in their chosen area of service. Feedback allows individuals to assume their identity as humans with individual scores, performances, abilities and characteristics. Feedback is the basis of how we feel about ourselves and we rely on it to continue shaping our ideas. The more of it we get, the better we are able to evaluate our abilities. In a society which deals in numbers, feedback is a chance to conserve our identity and enjoy our work. As such it should be at the heart of any training program.

RESULTS

OVERALL SIGNIFICANCE AT BETTER THAN .05 LEVEL.

REQUIRED 7.82 OBTAINED H = 8.19

AFTER THE CORRECTION Hc = 8.21

GROUP COMPARISONS

			H
I	vs II	LESS THAN .05 LEVEL -	2.51
I	vs III	HIGHER THAN .01 LEVEL-	10.15
I	vs IV	HIGHER THAN .01 LEVEL-	23.34

FIGURE 1

BIBLIOGRAPHY

1. CURTIS, F.D. and WOOD, G.G., A study of the relative teaching value of four common practices in correcting examination papers, School Review 1929, 37, 615-23

2. Hurlock, E.B., An evaluation of certain incentives used in school work, Journal of Educational Psychology, 1925, 16, 145-49

3. Page, E.B., Teacher comments and student performance; Journal of Educational Psychology, 1958, 49, 173-81

4. Plowman, L. and Stroud, J.B., Effects of informing pupils of the consequences of their responses to objective test questions. Journal of Educational Research, 1942, 36, 16-20

5. Sassenrath, J.M. and Garverick, C.M., Effects of differental feedback from examinations on retention and transfer. Journal of Educational Psychology, 1965, 56, 259-63

6. Stone, G.R. The training function of examinations. Research report No. AFPT RC-TN-55-8, 1955. USAF Personnel training research center, Lackland Air Force Base.

BEST AVAILABLE COPY

"PROBE": An automated testing system for the evaluation of ability and/or willingness to profit from a training situation

David B. Vinson, Ph.D.
1210 Medical Towers Bldg.
Houston Texas 77025

In 1970, COG, a computer software program capable of assessing the effect of cerebral concussion on cognitive function was reported. Using regression analysis, COG compared baseline data with post-injury performance data to calculate the probability of impaired cognitive function associated with cerebral concussion.

Building on experience gained with COG, a computer software program PROBE was developed to evaluate certain cognitive and personality functions, and from the interaction of these functions as specified by algorithms, to predict the subject's ability and/or willingness to profit from a training program.

In the PROBE system, performance data are acquired by the administration of tests to assess selective attention, short-term memory, rate of inductive reasoning, speed and accuracy of continuous addition, level of aspiration, ability to read, and mental attitudes. Certain biographical data are input to the system, but sex and culture are not computed.

Written in Fortran IV, the PROBE software uses approximately 27 K characters with two overlays of 11K and 9K respectively. Using an XDS-940 in time-share, performance data are input and compared with stored information to generate a report which quantitatively grades a subject on five factors previously identified by factor analysis as predictive of supervisor ratings of job performance.

The five factors as defined in the PROBE print-out reports are mental grasp, self-confidence, cooperativeness, emotional stability, and leadership. PROBE gradings are from 1.0 to 5.0, reported to one decimal point. A grading of 3.0 represents the mean of high performing normal subjects. PROBE has been on-line since 1971. From receipt of data, to user's receipt of the report, the turn around time is less than ten minutes per subject, anywhere in North America.

The system has the advantage of using non-professionals for the acquisition of biographical data and for the administration of the subtests of cognitive and personality functions. Most subtests are paper-and-pencil: the short-term memory subtest is automated. With the exception of one test, the Color Word test, the battery is appropriate for group administration and the subject records his own responses on an answer sheet. For the Color Word test, the administrator records the responses. Performance data may be transmitted to the service bureau for processing by any of several telecommunication modes.

The data base for the PROBE software is in the many thousands of subjects and includes the so-called normals (those who apply for jobs, demonstrate their competence by earning promotions, provide for their own needs, and for the needs of their dependents). To establish limits of performance, other classes of subjects were input to the data base: patients of known psychiatric or neurologic diagnoses, adolescents, the aged, stress-seekers, competitive athletes, and several classes of the so-called unemployable.

The PROBE report sample is shown as Figure 1. If desired, values of the five parameters can be graphically displayed (by hand or by X-Y plotter) as shown in the Starr-Plot, Figure 2. The plotting points of the Starr-Plot represent performance as it falls about the mean, where 3.0 represents the mean, 2.0 one standard deviation above the mean, 4.0 one standard deviation below the mean, and etc.

PROBE is designed to yield a report for use in work and/or training situations. From the same performance data which have been acquired for PROBE, a PDX report can be generated which is designed for use in medical evaluation or treatment situations. A PDX report, shown in Figure 3, is based on the same performance data used to generate the PROBE report in Figure 1.

Comparisons between predicted and observed ability and/or willingness to profit from a training program have been made extensively. Four formal validations of PROBE are summarized as Studies A through D.

STUDY A: Five professional football teams each tested seventy players. As the external criterion, players were graded by coaching staff members who had observed the player at least one year. By correlation analysis, these gradings were compared with PROBE gradings. (Although the language of the football reports is keyed to sport, the algorithms which predict ability and/or willingness to profit from a training program are essentially the same as those of PROBE.)

Although supervisor's ratings of work performance are generally acknowledged to be biased toward the mean, gradings of professional football coaches are not. Coaches' gradings of certain observable behavior associated with on-the-job performance are as shown in Table 1.

TABLE 1

TEAM	LEARNING	CONFIDENCE	COACHABILITY	AGGRESSIVENESS
A	.01	.01	.05	.01
B	.05	.10	.01	.01
C	.001	ns	.001	ns
D	.01	.10	.10	.05
E	.02	.05	.05	ns

STUDIES B-C: Performance data were acquired on 30 white (Study B) and 43 Negro (Study C) telephone operator trainees to test the null hypothesis that PROBE gradings would differ from supervisor ratings, observed behavior, and/or performance data to an extent no greater than chance.

Subjects to Studies B and C were totally unselected, that is, each applicant for work in a given time period was hired, tested and trained by the company unless she was excluded by medical findings of tuberculosis or venereal disease. Subjects were assessed by PROBE within hours of reporting to work for the first time. After each employee had been on the job for 90 days, each was subjectively rated by supervisors to whom PROBE reports were not available.

The company wished to evaluated on five parameters it felt were important in job performance: Knowledge of Work, Productivity, Pride in Work, Learning Rate, and Overall Performance. To test the relevance of the five PROBE parameters, the supervisors were first asked to rate the subjects on the company's five parameters and the five parameters of PROBE. For Study B, the agreement is as shown in Table 2.

TABLE 2

	Mental Grasp	Confidence	Cooperativeness	Stability	Leadership
Knowledge of Work	.001	.001	ns	.001	.001
Productivity	.001	.01	ns	.01	.001
Pride in Work	.001	.001	.001	.001	.001
Learning Rate	.001	.001	.05	.001	.001
Overall	.001	.001	.05	.001	.001

For Study C, the agreement is as shown in Table 3.

TABLE 3

	Mental Grasp	Confi-dence	Coopera-tiveness	Stab-ility	Leader-ship
Knowledge of	.001	ns	ns	.05	.001
Productivity	.001	ns	ns	ns	.001
Pride in Work	.02	.05	.001	.001	.001
Learning Rate	.001	ns	ns	ns	.001
Overall	.001	ns	ns	ns	.001

From the information in Tables 2 and 3, it appeared the supervisors (1) did find the five parameters of PROBE relevant to observed job performance: (2) saw the two races differently: and (3) graded "globally," that is, as seeing people as having much the same grading no matter what was being graded.

When PROBE report gradings were compared with supervisor ratings, the agreements are in line with those generally reported in the literature whenever any test performance is compared with supervisor gradings. The agreement between supervisor ratings of observed bheavior and/or performance and PROBE, for 30 white telephone operator trainees (Study B) is as shown in Table 4.

TABLE 4

	Mental Grasp	Confi-dence	Coopera-tiveness	Stab-ility	Leader-ship
Knowledge of	.01	ns	ns	ns	ns
Productivity	.01	ns	ns	ns	ns
Pride in Work	ns	ns	ns	ns	ns
Learning Rate	.001	ns	.10	ns	ns
Overall	.01	.01	ns	ns	ns

The agreement between supervisor ratings of observed behavior and/or performance and PROBE for 43 Negro telephone operator trainees (Study C) is as shown in Table 5.

TABLE 5

	Mental Grasp	Confi-dence	Coopera-tiveness	Stab-ility	Leader-ship
Knowledge of	.01	.05	ns	ns	.10
Productivity	.01	.05	ns	ns	.10
Pride in Work	ns	.001	.10	ns	ns
Learning Rate	ns	.10	ns	ns	.10
Overall	.10	.10	ns	ns	ns

STUDY D: Performance data were acquired from 855 employee candidates of U.S.Steel (Texas Works). On each subject, an early model of PROBE was processed. These raw and computed data were forwarded to the Department of Statistics, Texas A & M University. On the same subjects, U.S.Steel provided job and training data; U.S.Steel assured Texas A & M University that employee candidates were neither hired nor not-hired on the basis of PROBE reports. The two sets of data (PROBE and U.S.Steel) were kept entirely separate until they reached Texas A & M University for analysis. U.S.Steel data consisted of several ratings performed by supervisors, and certain data-of-record on hired/not hire, absenteeism, grievances filed, and etc.

A brief summary of the findings of the regression analyses of the data follows as Table 6.

TABLE 6

SUBJECTS	PROBE	PREDICTS U.S.STEEL RATINGS OF:
All	All parameters	Work performance. Keeps safety rules. Cooperativeness & conformity. Emotional stability. Authoritarian leadership. Industrial engineering rating.
All	Mental Grasp	Times changed departments.
All	Conformity	Cluster placement. Acceptance overtime. Acceptance assignments.
All	Confidence	Unrealistically high, poor cluster placement. Unrealistically high, more absenteeism. Unrealistically high, poorer work performance. Unrealistically high, poorer acceptance assignments.
Females	All parameters	Learning. Hired/Not Hired.
Females	Mental Grasp	Emotional stability. Cooperativeness & conformity.
Females	Conformity	Absenteeism.

SUBJECTS	PROBE	PREDICTS U.S.S. RATINGS OF:
Clerical-technical	All parameters	Hired/Not Hired. Non-authoritarian leadership.
Clerical-technical	Mental Grasp	Poorer PROBE, more likely hired· (men only).
Clerical-technical	Conformity	Safety violations
Clerical-technical	Cooperativeness	Absenteeism· better PROBE, more absenteeism.
Clerical-technical	Confidence	Safety violations.
Production, Latin	All parameters	Authoritarian leadership.
Production, Latin	Mental grasp	Poorer PROBE, better supervisor ratings of emotional stability.
Production, Latin	Confidence	Unrealistically high, poorer supervisor ratings of emotional stability.
Production, Latin	Competitiveness	Poorer PROBE, better supervisor ratings of emotional stability.
Production, Latin	Cooperativeness	Emotional stability.
Production, White	All parameters	Authoritarian leadership.
Production, White	Mental Grasp	Times changed departments.
Production, White	Cooperativeness	Emotional stability.
Production, White	Confidence	Unrealistically high, poor cluster placement. Unrealistically high, more absenteeism. Unrealistically high, poor work performance. Unrealistically high, more safety violations. Unrealistically high, poor acceptance assignments.
Production, Negro	All parameters	Emotional stability.
Production, Negro	Mental Grasp	Better PROBE, more acts of physical aggression. Acceptance overtime.

SUBJECTS	PROBE	PREDICTS U.S.S. RATINGS OF:
Production, Negro	Mental Grasp	Absenteeism. Better PROBE, more grievances filed.
Production, Negro	Cooperativeness	Better PROBE, poorer cluster placement.
Production, Negro	Conformity	Cluster placement.
Production, Negro	Competitiveness	Safety violations.
Production, Negro	Confidence	Unrealistically high, poorer cluster placement.

(In Table 6, all predictions are in the expected direction unless otherwise noted.)

SUMMARY: Although performance data for PROBE may be acquired by non-professionals, by the use of telecommunication nets, computer hardware and computer softwares, professional judgment and experience are used in evaluating those data.

PROBE is independent of the constraints of professional availability and cost, geography, personal bias of the examiner, and PROBE is designed to conserve time.

There is no practical limit to the number of persons who can be evaluated by PROBE in a fast, reliable, valid and cost-effective way, and this evaluation predicts ability and/or willingness to profit from a training situation. It is felt PROBE could be of significant usefulness in the evaluation of applicants for military duty and/or for the evaluation of those on military service.

By use of regression analysis, base-line performance data of PROBE can be compared with any subsequent retest performance data, and the probabilities of change can be calculated. PROBE may also be used in serial studies to determine reversibility of pathologic changes associated with aging, disease or trauma.

CHRONOLOGICAL
REFERENCES

1. 1935 Stroop, J.R.: Interference in serial verbal reactions. J. Exp. Psychol.

2. 1943 Lewin, K., Dembo, T., Festinger, L., & Sears, P.: Level of aspiration, IN: Personality and the Behavior Disorders, Editor J. McV. Hunt. New York: Ronald Press

3. 1948 Thurstone, L.L. & Thurstone, T.G.: SRA Primary Mental Abilities, Intermediate Form AM. Chicago: Science Research Associates.

4. 1949 Eysenck, H.J.: Schizothymia-Cyclothymia as a dimension of personality. Maudsley Hospital Report, London.

5. 1958 Eysenck, H.J.: Scientific Study of Personality. London: Routledge and Kegan Paul

6. 1960 Vinson, D.B.: Objectivity in the assessment of the thyrotoxic patient. J. Psychosomatic Research

7. 1960 Vinson, D.B. & Robbins, L.R.: Objective psychologic assessment of the throtoxic patient and the response to treatment. J. Cl. Endocrinology and Metabolism

8. 1960 Vinson, D.B. & Gaitz, C.M.: Objective measurement of psychobiologic decline. Proceedings, 5th Int. Congress Gerontology

9. 1961 Vinson, D.B.: Objectivity in the assessment of psychobiologic decline. Vita Humana

10. 1963 Vinson, D.B., Barnes, T.G., Russell, H.E.: The effect of altered consciousness on information processing. Schellenger Res. Labs.

11. 1964 Smith, G.J.W. & Nyman, G.E. The serial colour-word test. A summary of results. Sthlm. Skandinaviska Testforl.

12. 1967 Vinson, D.B.: Information processing in the human subsystem. Instrument Soc. of America (Apollo section)

13. 1967 Vinson, D.B.: Information processing in man-machine systems. J. Assn. Adv. Medical Instrumentation

14. 1967 Vinson, D.B.: A study of the effect of time on mental efficiency and emotional stability. Am. Psychol. Assn. Annual Meeting, Div. 20, Washington DC

15. 1969 Vinson, D.B.: Performance testing as it relates to injury. Nat'l. Research Council, Division on Medical Sciences

16. 1970 Vinson, D.B.: Computer evaluation relative to injuries. 12th Nat'l. Conf., Medical Aspects of Sports, Boston

17. 1971 Vinson, D.B., Constant, G.A., Quackenbush, J.R. & Molloy, J.P.: Automated mental status examination: pre and post drug effects. Drug Information Assn., Computer Workshop, Chicago

18. 1972 Vinson, D.B.: Automated assessment of changes in human performance. Grand Rounds, Baylor University College of Medicine, January 21.

19. 1972 Vinson, D.B.: Telecommunications and the automated mental status examination. Nat'l. Telecommunications Conference, Houston.

20. 1973 Vinson, D.B.: Can computer programming predict post-operative return to function? Faculty, "What's New in Orthopaedics," Am. Acad. Orthopaedic Surgeons, San Antonio

21. 1973 Vinson, D.B. & Harper, C.R.: PRIME: a multi-language automated psychiatric screening tool. Civil Aviation Medical Assn., Guadalajara, Mexico

FIGURE 1

ASSESSMENT SYSTEMS, INCORPORATED
"PROPE"
OCTOBER 12 1973

ASI-123456789

MENTAL GRASP: 3.1 THE ABILITY TO LEARN FROM A SYMBOLIC REPRESENTATION OF A PROCESS OR PROCEDURE. THE ABILITY TO UNDERSTAND HOW A SUB-SYSTEM RELATES TO A TOTAL SYSTEM. THE ABILITY TO THINK OF ALTERNATE SOLUTIONS TO PROBLEMS.

SELF CONFIDENCE: 2.4 THE INDIVIDUAL'S FEELING OR EXPECTATION THAT HE WILL MASTER A PROCEDURE AND/OR PERFORM AS SUCCESSFULLY IN FUTURE AS HE HAS IN THE PAST.

COOPERATIVENESS: 2.0 IDENTIFIES WITH AUTHORITY. TAKES DIRECTION FROM THOSE IN AUTHORITY. IDENTIFIES WITH CORPORATE GOALS. SUBORDINATES PERSONAL GOALS TO CORPORATE GOALS.

EMOTIONAL STABILITY: 4.0 CONSIDERS THE OUTCOME BEFORE HE ACTS. IS ABLE TO CONTROL TEMPER AND TO TOLERATE FRUSTRATION. SETS LONG-TERM RATHER THAN SHORT-TERM GOALS FOR HIMSELF.

LEADERSHIP: 2.9 THE ABILITY TO TAKE COMMAND, TO INSPIRE CONFIDENCE AND TO ENLIST THE COOPERATION OF OTHERS.

IN EACH CATEGORY A RATING OF
1 = SIGNIFICANTLY ABOVE AVERAGE
2 = ABOVE AVERAGE
3 = WITHIN NORMAL LIMITS
4 = BELOW AVERAGE
5 = SIGNIFICANTLY BELOW AVERAGE

THIS ANALYSIS MAY REFLECT THE SUBJECT'S PHYSICAL AND EMOTIONAL STATUS AT THE TIME OF TESTING AND SHOULD NOT BE USED AS A SOLE CRITERION FOR ACTION OR DECISION. THIS REPORT WILL BE SIGNED ONLY WHEN TEST FINDINGS, SCORING AND CALCULATIONS ON WHICH REPORT IS BASED HAVE BEEN VERIFIED.

----------PH.D.

FIGURE 2

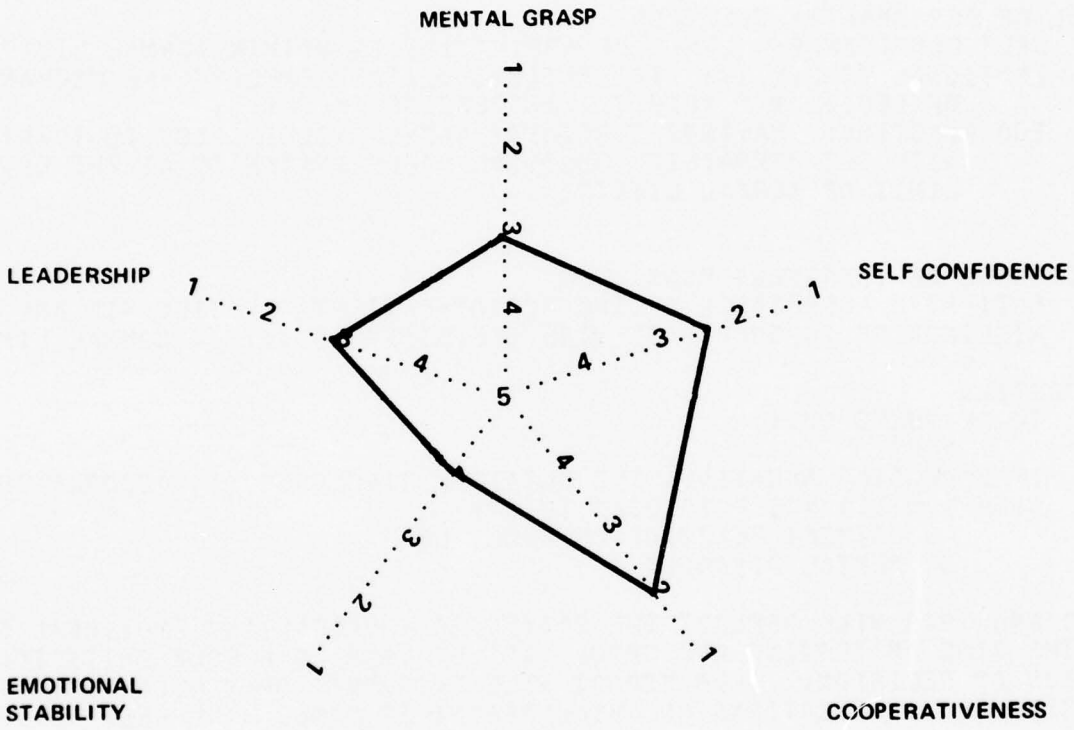
Assessment Systems, Inc.

CNO: ASI-123456789

DATE: October 12 1973

FORMAT: "PROBE"

STARR PLOT



(c) H.C. VINSON, 1970

FIGURE 3

ASSESSMENT SYSTEMS, INCORPORATED
"PDX"
OCTOBER 12 1973

ASI-123456789

LEVEL OF COGNITIVE FUNCTION
ATTENTION IS AT THE UPPER LIMIT OF NORMAL LIMITS.
CONCEPTUAL THINKING: ABILITY TO ISOLATE A COMMON CHARACTERISTIC AND TO SHIFT FROM ONE PRINCIPLE OF CLASSIFICATION TO ANOTHER IS AT THE LOWER LIMIT OF NORMAL LIMITS.

LEVEL OF PERSONALITY FUNCTION
SELF CONFIDENCE: LEVEL OF ASPIRATION IS WITHIN NORMAL LIMITS.
EMOTIONAL STABILITY: THE PATIENT NEITHER EMPLOYS THE MECHANISM OF DENIAL NOR ACTS OUT AGGRESSIVE IMPULSES.
EGO FUNCTION: PATIENT'S ABILITY AND/OR WILLINGNESS TO IDENTIFY WITH THE PREVAILING CULTURAL VALUE SYSTEM IS AT THE LOWER LIMIT OF NORMAL LIMITS.

ACCEPTANCE OF TREATMENT PLAN
PATIENT'S ACCEPTANCE OF THE DOCTOR-PATIENT RELATIONSHIP AND WILLINGNESS TO COOPERATE WITH THE STAFF IS WITHIN NORMAL LIMITS.

IMPRESSION
TO BE RULED OUT:
NONE
IF FOREGOING NEGATIVE, THE CLINICAL DIAGNOSES (IN ACCORDANCE WITH DSM-II) ARE PREDICTED TO BE:
ADJUSTMENT REACTION OF ADULT LIFE
NO MENTAL DISORDER

THIS ANALYSIS WILL REFLECT THE PATIENT'S PHYSICAL AND EMOTIONAL STATUS AT THE TIME OF TESTING AND SHOULD NOT BE USED AS A SOLE CRITERION FOR ACTION OR DECISION. THIS REPORT WILL BE SIGNED ONLY WHEN TEST FINDINGS, SCORING AND CALCULATIONS ON WHICH REPORT IS BASED HAVE BEEN VERIFIED.

----------PH.D.

Prospective Chief Petty Officer Leadership/Management Seminar

Commander G. C. Hinson
United States Coast Guard Training Center

INTRODUCTION

The need for leadership in the military services is axiomatic. At all levels, managers--those with a responsibility to get a job done--must be able to direct other men to want to do a job well. Recently, within the Coast Guard, there has been increased emphasis placed on the leadership abilities required of E-7 through E-9 Chief Petty Officers. Late in 1972, the Commandant tasked the Training Center at Governors Island, N. Y. with defining the elements of leadership a Chief Petty Officer must possess, and preparing a "nuts and bolts" course to teach these prospective leaders this managerial skill. This paper outlines the procedures that were used in development of such a course.

BACKGROUND

What is the essence of leadership? Can it be taught in a classroom? Can this be taught at all? Can it be taught to everyone? Can men be transformed into leaders? A Questionnaire sent to E-9 Chief Petty Officers revealed a leader to be intelligent, a specialist in his rate, imaginative, dependable, capable of making quick decisions, especially under pressure, and additionally one who inspires his men. Could we develop a course to produce such a person? Is this the man who could get the job done? Research in this field showed that many leadership courses are available. Most courses are designed to shape men into what someone thinks leaders ought to be. In general many efforts have been disappointing and a waste of time and money. Even extensive training at military academies has failed on many occasions to transform cadets into effective leaders. The reasons for our apparent failures should be explored, are our methods of instruction, the students or a combination of these two factors to blame; or is something else at fault? What should be changed to improve leadership training? Many studies show that the most important dimension of a leader is his relationship to his subordinates. He will be most effective when his men respect and trust him. But respect and trust are built upon something deeper than personality and friendship. This made a lot of sense when we re-read the characteristics cited on the CPO questionnaire: intelligent, specialist in rate, imaginative, dependable, capable of making a quick decision, especially under pressure. It is clear that respect and trust must be resultants of performance oriented characteristics. Fred E. Fiedler has conducted extensive research on leadership at the University of Illinois. His studies show that it makes little sense to speak of a good or poor leader. There are only leaders who perform well in one situation but not in another. In these situations, he differentiates between task-motivated and relationship-motivated leaders. Those primarily

interested in building personal relationships with their fellow workers are said to be relationship oriented; those interested primarily in getting the job done are considered task motivated. Fiedler also asserts that a task-motivated leader will have a greater probability of success than a relationship motivated leader. Based upon our analysis of the literature available, a course in leadership should stress (1) the psychology of motivation and the satisfaction of doing a job well, (2) analysis and control of leadership situations, and (3) building a relationship of mutual respect between the leader and members of the group. This was the basis for the Coast Guard Leadership Course and our challenge was to develop realistic rather than idealistic leaders.

HISTORY OF DEVELOPING THE COURSE

E-7 through E-9 Chief Petty Officers are Coast Guard front-line managers. They most frequently fill middle management roles serving to clear the way for command actions and letting the command know how the crew feels. But on the other hand, Chiefs are assigned smaller units as Officer-in-Charge and as Executive Officer. These possible assignments require having both a narrow specialized knowledge in a technical field and broad knowledge of the service in general. They also must be able to communicate vertically and horizontally within the command. Our leadership course was tailored then to meet these needs. We sought also to provide prospective Chiefs with a basis for decision making. A leader has three methods of exercising his skill: he can be authoritarian, democratic or depend upon his men to exercise responsibility and good judgment to get things done. Although the good leader must use all three techniques to effectively operate, current literature promotes democratic leadership or group dynamics as the preferable choice. This conflict results in confusion. To quote the Harvard Business Review:

> "often he [the leader] is not quite sure how to behave; there are times when he is torn between exerting 'strong' leadership and 'permissive' leadership. Sometimes new knowledge pushes him in one direction ("I should really get the group to help make this decision"), but at the same time his experience pushes him in another direction ("I really understand the problem better than the group and therefore I should make the decision"). He is not sure when a group decision is really appropriate or when holding a staff meeting serves merely as a device for avoiding his own decision making responsibility."

We felt a Chief's decision making ability would be proportional to the information he had to draw upon. Consequently, we stress the forces operating in the decision making process.

(1) fiscal restraints
(2) responsibility and authority
(3) legal boundries

(4) military considerations

(a) chain of command
(b) personnel evaluation
(c) discipline

The successful leader is aware of all relevant forces and can put them into perspective when making a decision. He is flexible enough to not fear change, when change may be indicated. Therefore, the Coast Guard Leadership Course emphasizes analysing situations, knowing available options, and selection of efficient strategy. Our course:

1. Supplies information (or indicates where information can be found) concerning relevant forces.

2. Teaches analysis techniques to enable one to know self, subordinates, superiors, and peers.

3. Stresses the selection of the most appropriate and productive leadership techniques, and

4. Provides the opportunity to develop communication skills.

SUMMARY

Our goal at the Training Center is to produce a realistic leader; one who is well grounded in Coast Guard programs, precedures, policies, and problems; one who knows his importance in the management of resources; one who is sensitive to the needs of his subordinates and his superiors. We feel that our four week Leadership Course can achieve these goals.

APPENDIX

Outline of Curriculum

Enclosure (1) to CO TRACEN Ltr 1510 dtd

PROPOSED PROSPECTIVE CHIEF PETTY OFFICER LEADERSHIP/MANAGEMENT SEMINAR

FIRST WEEK

Monday

a. Introduction to the training center and check in.
b. Greeting by Commanding Officer, Executive Officer, Senior Chaplain, and the President of Governors Island Chief Petty Officers Association.
c. Explanation of the seminar's methods and goals - emphasizing skill, not theory.
d. Administer the Leadership Opinion Questionaire
e. Present historical leadership studies.

Tuesday

a. In depth course orientation; How to accomplish the job
b. The psychology of motivation including perception, role identification, communications and role playing.
c. Achievement motivation including security vs. challenge quiz and review (quizzes are personality preference measures).

Wednesday

a. Continuation of achievement motivation including Machiavellianism and summary.
b. Leadership styles, Fiedler Development.
c. Leadership Matrix and review Leadership Opinion Questionaire questions.

Thursday

a. Situation evaluation and solution choosing - obtaining facts prior to decision making.
b. Return LOQ and discuss the results.
c. Leadership games, including "Survival on the Moon".

Friday

a. The importance of communicating, encouraging suggestions.
b. Retake and review of LOQ.
c. Review and discuss pain research.

SECOND WEEK

Monday

a. The Coast Guard, including the movie "The Eighth Mission", Coast Guard Organization and the Chain of Command.
b. Career Counselling which includes methods of counselling, self understanding, human understanding, evaluation of people, basic salesmanship, effective listening, counselling problems, interviewing and specific areas of CG personnel policies (rotation, etc), V.A. assistance, social security, insurance survivors benefits and annuities, personal financial management, civilian careers both after retirement and the current job market.

Tuesday

a. Drug abuse including types of drugs in use and the reasons people turn to drugs.
b. Continuation of career counselling.

Wednesday

a. Drug abuse continued; alcohol, and the Drug Exemption Program.
b. Discipline and grievances.
c. The Uniform Code of Military Justice.

Thursday

a. Human Relations
b. Body Language
c. Fundamentals of recruiting

Friday

a. Human Relations
b. Public speaking
c. Public information

THIRD WEEK

Monday

a. Human relations
b. Budget process including Sub-Head 30, planned projects and procurement, cost consciousness
c. Allowance lists
d. Procurement, including advance planning, work lists vs. procurement lead time.

Tuesday

a. Human relations
b. The Personnel Manual (CG-207)

Wednesday

a. Human relations
b. Title "B" property
c. Personnel records
d. Health records

Thursday

a. Pay and Allowances - pay office proceedures
b. Travel - household effects
c. Leave, leave rations, computation, types
d. CHAMPUS benefits, elegibility, orthodontics
e. Medical
f. Retirement

Friday

a. Educational opportunities, including service schools, CGI, USAFI and off duty education.
b. Advancements, including career patterns, servicewide exams.
c. uniform regulations.
d. training procedures.

FOURTH WEEK

Monday

a. Organization, delegation and planning of relatively complex tasks, including long range goals and intermediate objectives and the need for follow up.
b. Clerical, including effective writing, correspondence manual.

Tuesday

a. Forms, reports and publications.
b. Classified material, handling stowage, access and the need to know.

Wednesday

a. Records keeping
b. Rules of leadership.

Thursday

a. CPO responsibility up and down the line: a general summation of what a CPO is, ranging from accident prevention through experience to rumor control by not participating in speculation; a chief's presence in the work area not the CPO mess; supporting the service and superiors and not adding to gripe sessions with junior enlisteds.

Friday

a. check out
b. Graduation - CO or XO address class - theme "Personal Integrity and Ethics".

ADDITIONAL TRAINING AS DESIRED (NIGHTS)

1. Self Taught Reading Improvement
2. Self Taught Mathematics Improvement
3. Self Taught Spelling Improvement

TENTATIVE SEMINAR TEXTS

1. "What Every Supervisor Should Know", Lester R. Bittle, Second Edition, 1969, McGraw-Hill Inc.
2. Naval Leadership, US Naval Institute
3. "Right Down the Line", Charles A. Pearce, Ed, 1955, Arrow Books, Inc.
4. Various articles on motivation, achievement and leadership.

TRANSCENDENTAL MEDITATION (TM): A THREAT OR TECHNIQUE IN FUTURE PERSONNEL EVALUATION

R. O. Waldkoetter
Consultant, Educational & Personnel Systems
Indianapolis, Indiana

How Can Personnel Evaluation For Human Resources Be Humanly Oriented?

The wide expanse that is perceived when we begin to think of ways to promote the best uses of human resources includes so much for attention, we may be disturbed by the complexity of this challenge. The human subject, worker or friend is not easily directed in the most predictable channels without occasionally interfering with our programed goals by time, territorial or value differences. In behavioral modification, training, and other conditioning experiences, a path is sought which can give desirable ethical and productive results without the majority of persons involved resisting the direction in which managed or compelled to follow.

Over the years in the endeavors addressed under the quest for adequate personnel evaluation techniques, there has been much concern about evaluation which can reasonably assure that a person with acceptable skill is selected for whatever opportunites found available. This goal has not diminished as greater and greater effort has been made to get at skill, ability, interest and personality factors which are believed to be fundamental sources of variance affecting behavior. As a position guiding measurement design, a creeping doubt is frequently dismissed whenever the problem of motivation is posed or the "sense of timing" awareness shown, which both seem to influence how skilled behavior is brought to bear in any problem situation.

With some assumptions about what have been generating forces for personnel evaluation -- measuring individuals primarily with paper-pencil tests in achieving some type of criterion behavior -- there remains also a principal question regarding the desire and persistence needed to maintain behavior after the record of achievement and aptitude is indexed. Whether we view persons measured, getting ready to be measured, or at a specified interval after measurement, a constraint is still present that makes one feel uneasy to the degree that a strong level of continuing motivation is not foreseeable unless very high qualifying standards are somehow applied. What choice is there for preparing independent behavior that will directly contribute to greater efficiency and eventual productivity without relying upon passive and coercive methods and authoritarian command?

If you will, let us suppose there are techniques by which we can heighten performance by improved attention, integrated responses, and less

particularly ego-centered goals. Under favorable conditions individuals will expend their intellectual and emotional resources to attain specified goals without being moved by harsh instructions and built in punishment should they not comply to set schedules. Again, if you will suspend critical analysis for a moment, I ask that you give your attention to the concept that the most convincing teaching and experience results from your having permitted yourself to be so engrossed in a practice task, your sense of awareness is almost wholly guided by the culmination of the action being performed. You seek only to determine the degree of execution by which you might achieve a finer shade of performance and in some way allow yourself to evaluate the results in terms of some other instructive experience. When the activity does not satisfy you with the anticipated results some corrective feedback is usually applied to obtain adjusted future responses toward the goal selected. Seldom did critical insight occur where one could ascertain that raw energy was not the primary motivating force, but rather with the proper amount of energy, coordinated and well-intergrated behavior most often furnished a wise problem-solving approach.

Can an Alternative Be: Human Awareness to Support Measurement?

You must recognize my introductory portion of this paper is setting the stage for a statement of my purpose and tentative problem resolution. Which is to request that you reflect on better means by which personnel evaluation, clinical-or objectively-orriented, can show some humanistic emphasis on an individual's state of awareness to identify the individual's locus of psychological control, before responding to outer-directed commands to reach given task standards of performance. One way to encourage intrinsic and humanistic motivation to perform up to the individual's potential is to maximize functioning of the overall neurological system. Although intensive classic conditioning and hypnosis or autosuggestion are potentially within the sphere of utilization, I do with all candor submit that the personalized locus of control as advanced by the technique of Transcendental Meditation, and its possible variants, offers the type of process through which learning and training can be selectively accelerated and nurtured.

Let's Look at a Technique to Strengthen Performance.

Transcendental Meditation (TM) is presented to you for a futuristic look, and is probably only one of several avenues of treatment to be utilized to improve the serenity of consciousness so that we may perform more efficiently. It is defined as a process wherein the individual sits usually in an upright position, as comfortably as possible, eyes closed, effortlessly lets the mind drift toward a mantra that was assigned by a teacher and initiator (Fiske, 1972). I would not try to even give a comprehensive definition of mind, other than to say it comprises that conscious awareness of the manifold fields of activities being mediated by the functioning of the brain and nervous system. The mantra, as mentioned, is a sound which sets up some sort of commanding password that seems to

shift the level of consciousness to let go of value-laden thought. Although not quite in the same way as hypnosis does, TM produces with the mantra applied, an almost immediate change in the mental and physiological functioning (Wallace, 1970).

This short definition is practically all that need be stated about the identity of TM as a technique. The agreement on its potential effect in personnel evaluation is yet to be acknowledged. So let us exercise the collective imagination to the point that we may admit there are many contingencies which will become evident in future developments in personnel evaluation, testing and personnel system analysis. Riding the crest of the wave with statistical applications and computerized innovations, testable behavior has been mainly explained as a dependent variable in response to tasks set before the human subject. Behavior not directly perceptible through the nearly always observable channels of psychomotor evaluation is routinely dismissed as not in the territorial interest of the practicing behavioral scientist. New models of human behavior should be proposed to account for altered states of consciousness and connected behavior. How may these states tend to produce systematic variations in behavior when all possible characteristics in selected individuals appear to be similar?

What Related Ideas and Functions of TM Can Be Suggested?

To a certain extent I believe Mc Clelland (1973) was speaking of something bordering on a change in a theory of testing when he specified "tests should involve operant as well as respondent behavior," and "tests should sample operant thought patterns to get maximum generalizability to various action outcomes." He goes on to say that "the tester of the future is likely to get farther in finding generalizable competencies of characteristics across life outcomes if he starts by focusing on thought patterns rather than by trying to infer what thoughts must be behind the clusters of action that come out in various factors in the traditional trait analysis." Yet Mc Clelland has only seen fit to approach the edges of the testing domain, and balances precariously to not imply further that we might well study those psychic variables which impact in the levels of consciousness and thus affect the thought patterns.

To get behind thought patterns for testing purposes will give interested scientists no end of quizzical moments when they attempt to learn about the "conditions of knowing" that many of us were taught to be forbidden territory. Today observing human energy research trends in biophysics and such interdisciplinary fields, if the psychophysicist does not emerge to work with the innovative experimental clinican the future of human understanding and measurement could well be removed from the hands of conservative theorists and practitioners. How could this ever happen experienced psychologists will say attempting to think of flaws in the logic of this paper. Very well, then, I will hazard several predictive views under which, the TM influence of defined levels of consciousness might give differing insights in relation to how certain varieties of measured behavior could be interpreted. Then, if these predictive views do

not suggest to you feasible study problems, I am mistakenly following heretical research developments, and personnel evaluation should be declared only a field for refining proven psychometric methodology.

Viewing the TM technique as a supportive treatment, let us visualize a situation where the measured output of an individual has become erratic and this person seems to be chronically fatigued. Unless there is some physiological damage, he should regain any loss of applied skill where normal loss is not expected to accompany a heavy degradation of skill. By using TM, Wallace (1970) has with advanced meditators, acting as their own controls, showed changes in EEG (more frontal alpha and theta waves), and decreased oxygen consumption, heart rate, cardiac output, and blood lactate, with increased skin resistance, and certain restful changes greater than many hours of sleep. The acquisition and recall of the individual's lost skill should proceed more swiftly as the refreshed subject or incumbent voluntarily experiences TM and makes an inte grated learning reaction after the psychophysiological adjustment induced by TM. One small pilot study (Abrams, 1972) has shown support for a direct relationship between TM and quicker acquisition and higher recall performance on a paired-associate learning and recall task. A smoother response to achieve certain performances could well be anticipated then as the individual self-directs his own quiet conservation of personal mental, physical, and emotional resources.

Another application of TM could be viewed in contrasting this technique with preparation and conditioning for intensive problem-solving tasks. In training and education systems, where severe experiences tax mental and physical endurance, shifting levels of consciousness might furnish that added increment of renewed vigor and need for achievement. How often is frustration brought about in learning, and defended as the need to "overcome resistance", only to induce inhibiting effects which extinguish desired learning. While test-taking anxiety has been significantly lowered in a study (Doucette, 1972), other reports have indicated decreases in negative personality characteristics and expansion of awareness described as self-actualization (Kanellakos and Bellin, 1972). Although highly conditioned responses to given tasks are required in critical behavioral sequences, the ability to adapt to varying circumstances and threats to composure could be enhanced by TM training. Finding how one goes about encouraging the trainee and job incumbent to accept readily the potential of such a technique for restoring and recreating vital energy does remain a lesson to be learned in interfacing individual and group dynamics.

As a different research viewpoint to establish what uses suggestibility may provide in learning and sustaining job standards, may we not attempt to distinguish the states of consciousness described by hypnosis and TM. And try to examine any probable similarities or unique features when these techniques are introduced into a learning enviroment. Repetition and conditioning practices are normally justified in training and management, as essential to sustain task achievement, and the individual is disciplined to infer that motivational pressures are largely unavoidable. By assigning carefully regulated and voluntary treatments with hypnotic learning sequences, the human being may adapt to situations in which he can monitor

his own hypnotic condition and complement his further awareness by the practice of TM. A heightened sense of human understanding is hypothetically in reach and the behavioral scientist would be ready to analyze aspects of consciousness all but left to imagination and theological speculation. This exploration of inner space, as with outer space, will not appeal to everyone and to some conjures up threats to sanity and human integrity. To motivate selected individuals to undergo experiences which to date have generally proven enriching should add excitement to learning while maximizing performance and creativity.

The person who risks his own well being and those material things acquired in the fully conscious realm of "getting and doing" will likely reconsider goals and values when exposed to different levels of consciousness in which spontaneous trial and error behavior are minimized. No tendency to coerce learning and task behavior should be tolerated, but when selected individuals master the experience of altering their conscious levels of awareness, they will contribute to greater creativity and composure in their living and working environments. Both political and religious movements have long profited from having dedicated individuals and a structured method for preparing themselves to face challenges as potential opportunities. In effect to know what conscious variables in altered states of awareness open a subject to greater insight for accelerating learned behavior, I believe, will reveal human resources only seldom tapped and hardly ever articulated as hopeful attributes for personnel evaluation.

Will you accept the existence of such a technique as TM and does it threaten or serve as an optimistic note in searching for well-integrated measures to predict the further exploration of promising human resources?

Can We Afford to Deny Consciousness in Personnel Evaluation?

Whenever a projected discussion of something like "states of consciousness" is put into written words, the fear that much of what is said will seem beyond rational comprehension begins to be very constraining. On the other hand the scientific endeavor to plow a new section of thought is even more provoking. Whether my discourse on TM will pique your interest or disdain, really should not matter. The basic purpose is to let a newer theoretical and experimental alternative be expressed so that personnel evaluation does not become too self satisfied with fundamental principles of human performance and the origin of measurable qualities.

Surely, operant behavior can be interpreted as resulting from the type of environment the individual gains through sensation and genetic inheritance. However, that environmental awareness is able to be altered by variables which are not immediately perceptible. Everything from the electromagnetic fields and atmospheric changes to psychic interference by adverse energy patterns developed between individuals seem to give us pause to wonder what does cause behavior we try to evaluate. In a recent publication Rogers (1973) made an enlightening review of some new challenges in psychology and pointedly questioned "is this the only reality." You need not

agree with his own brand of clinical methodology, yet I feel great significance must be put in his serious query about needing to risk the investigation of paranormal phenomena.

An approach as seen through the channel of TM can help open the dark door so prohibited by conservative investigators when inquiring about the other reality. Where human beings in all walks of life have carelessly sought a finer awareness with drugs and other illusory substances, that behavior has led to many unfortunate incidents. At least two Army general officers have advocated voluntarily introduced presentation of TM for interested personnel, but only after being initiated into TM practice themselves and experiencing personally favorable results (Eastman, 1973).

New horizons are going to be attained as uses of TM and other states of consciousness are delved into to discover the sources and functioning of our thought patterns. If we encounter new experiences that are self actualizing for us and lead to finer harmony within and beyond our own degrees of awareness, we can be more confident in understanding individual behavior we are assessing and how human resources should be properly used in diversified organizational climates. There will be a constant urge to study the complete human system in the interaction with environment and productive results being obtained. The emphasis on methodology in evaluation cannot be ignored, however the finest analysis procedures are misguiding if we are not evaluating the skills or traits we think should be under study.

In searching for various ways to measure and apply our findings about human behavior, the goal to conserve and develop a full range of human resources must not be limited by lack of newer concepts and subsequently methods in personnel evaluation. During a recent lecture by a well known astronaut (Mitchell, 1973), he stated that he was convinced of the need to study the spectrum of consciousness displayed by the human organism, and has organized such a group for this purpose. With this kind of forceful inquiry taking place in numerous international study centers, we must agree such techniques as TM should be of some interest in our professional frame-of-reference. There are human abilities which apparently transcend the traditional explanation of how skills are normally acquired and used. We should see what effects these abilities have on our theories and evaluation practices even if the prospect seems unsettling. In the hope that my discussion of Transcendental Meditation has not alienated your own creative urges, I end with this meditative remark attributed to Einstein: "... Imagination is more important than knowledge."

REFERENCES

Abrams, A. I., Paired associate learning and recall: a pilot study comparing transcendental meditators with non-meditators. Unpublished progress report. Berkeley: University of California, Education Department, February, 1972.

Doucette, L. C., Anxiety and transcendental meditation as an anxiety reducing agent. Unpublished paper. Hamilton, Canada: Mc Master University, January, 1972.

Eastman, M., The military meditators. Army/Navy/Air Force Times. July 4, 1973.

Fiske, E. B., Thousands finding meditation eases stress. The New York Times, December 11, 1972.

Kanellakos, D. P. and Bellin, W., The practice of meditation as a means to the fulfillment of the ideals of humanistic and transpersonal psychology. Paper presented at the 10th Annual Meeting of the Association of Humanistic Psychology, Honolulu, Hawaii, August, 1972.

Mc Clelland, D. C., Testing for competence rather than for "intelligence". American Psychologist, 1973, 28, 1-14.

Mitchell, E. D., Outer space and ESP. *Lecture presented* at the 3rd Annual PSI, Inc. Conference, Indianapolis, Indiana, October, 1973.

Rogers, C. R., Some new challenges. American Psychologist, 1973, 28, 379-387.

Wallace, R. K., Physiological effects of transcendental meditation: a proposed fourth state of consciousness. Unpublished doctoral dissertation. Los Angeles: University of California, 1970.

"IMPLICATIONS OF CARREL INSTRUCTION ON MILITARY TESTING" A PRENARRATED SLIDE PRESENTATION BY DR RONALD W. SPANGENBERG OF THE MEDIA DESIGN & EVALUATION BRANCH, TECHNICAL TRAINING DIVISION, AIR FORCE HUMAN RESOURCES LABORATORY

AFHRL TECHNICAL TRAINING DIV.
MEDIA DESIGN & EVALUATION BRANCH

Focus frame

1

This presentation is entitled Implications of Carrel Instruction on Military Testing, Ron Spangenberg speaking. For proper synchronization you should now be seeing the specially designed focus frame of the Media Design & Evaluation Branch of the Technical Training Division, Air Force Human Resources Laboratory.

CARREL INSTRUCTION

Student working at standard carrel

2

What does carrel instruction mean to you?

Basic Electrical and Electronics School, U.S. Naval Training Center, San Diego, California

Students using workbooks and programmed texts at "dry" carrels. Some test equipment on shelf.

3

For the Navy student at San Diego it means a programmed instruction booklet with some items of electronic test equipment.

Individual Learning Center
U.S. Army Infantry School
Ft Benning, Georgia

Study carrel containing:
35mm slide projector, 3/4" V-matic format video cassette player, Super -8 film cartridge device; 12" color video monitor, synch. audio cassette deck

4

For the planners at Ft Benning, carrel instruction means media.

CARREL INSTRUCTION - SPANGENBERG
AUDIO-VISUAL STORY BOARD AFHRL/TT

College Library
Mt San Jacinto College
Gilman Hot Springs, California

Study carrel containing: 35mm filmstrip projector and reel to reel tape playback unit.

5

The carrels at Mt San Jacinto College have prenarrated filmstrips.

UPT Learning Center
Williams AFB, Arizona

Two-man study carrel designed to represent T37 cockpit. Audio selection directs the student through procedural lessons.

6

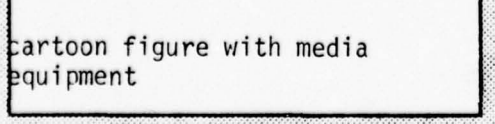
Undergraduate pilot trainees at Williams work with A/C control panels

Student watching Audiscan with Security Police Training Program

7

while Security Police at Lackland can view audiscan presentations at a table.

MEDIA DESIGN & EVALUATION LAB
AFHRL



8

In the Air Force Human Resources Lab the Media Design and Evaluation Branch of the Technical Training Division has been involved in

CARREL DESIGN Carrel with calculator and tape recorder visible. How to operate calculator program is on screen.	carrel design and instruction for several years.
9	
Weapons Mechanics Course Compulsory Remedial Training Lowry AFB, Colorado Students viewing prenarrated slide presentations at standard carrels.	We have set up carrels in a learning center presenting prenarrated slides most frequently.
10	
Weapons Mechanics Course Compulsory Remedial Training Lowry AFB, Colorado Student viewing PSM-6 Program at Performance Aiding Carrel.	We have designed and built a performance aiding carrel in which an inset panel in the center of the working surface can be replaced with specific performance boards.
11	
Panel inset for Performance Aiding Carrel with wiring for making continuity checks.	Such as this board designed to refresh the students in making continuity checks with the PSM 6 Multi-meter.
12	

AUTHORING

Carrel with two Carousels, 8mm projector and audio playback unit. Random access slide selector is shown.

13

Three conclusions we have reached are: (1) carrel instructional materials must be developed, produced and validated locally,

INTERACTIVE CARREL

Interactive carrel with two persons viewing an Audiscan presentation

14

(2) carrel instruction is not necessarily performed in isolation. We feel students frequently should work together as in this interactive carrel which is designed to help two students work together - or for an instructor to help a student,

The Audio Visual Instructional Display System module being demonstrated by a secretary

15

(3) the third conclusion is that a carrel is not a piece of library furniture. This portable carrel can be rolled to actual pieces of equipment and provide hands-on instruction. To us a carrel could consist simply of a set of earphones.

CARREL INSTRUCTION

Student working at standard carrel.

16

You can conclude carrel instruction may be many different events. However, there are three areas of evaluation which will receive increased emphasis due to the use of carrel instruction.

CARREL INSTRUCTION

QUALITY CONTROL

Student working at standard carrel.

17

These three areas are the quality control of student performance

CARREL INSTRUCTION

QUALITY CONTROL

MODULE VALIDATION

Student working at standard carrel.

18

the validation of modules of mediated instruction

CARREL INSTRUCTION

QUALITY CONTROL

MODULE VALIDATION

LEARNING EFFECTIVENESS

Student working at standard carrel.

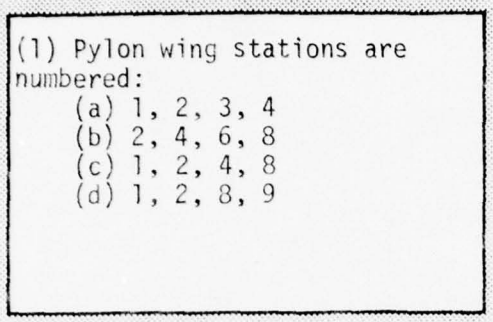
19

and using evaluation techniques to increase learning effectiveness. One of the first implications for testing in the area of learning effectiveness

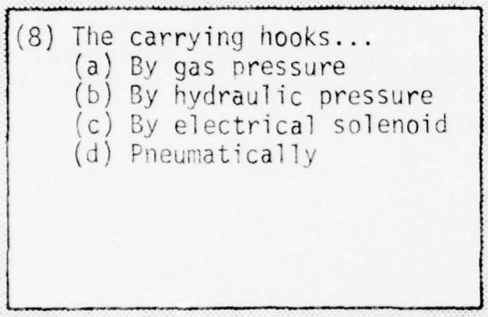
DIAGNOSTICS

Student pressing button on Mark II Autotutor.

will be an increase in the development and use of diagnostic tests to direct the learning of the student. The student the instructor or the program will use diagnostic tests to select modules for optimal paths of learning.
Frank Graham, an instructor in the Weapons Mechanic Course at Lowry, developed some prenarrated sli modules for compulsory remedial training. As p of each module he included slides such as these.

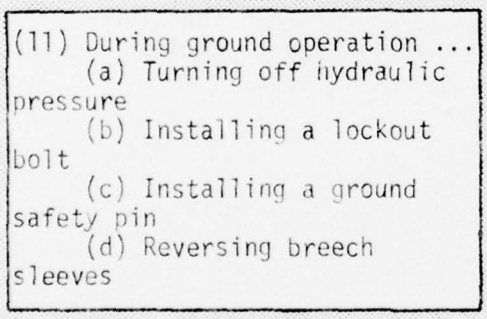


21

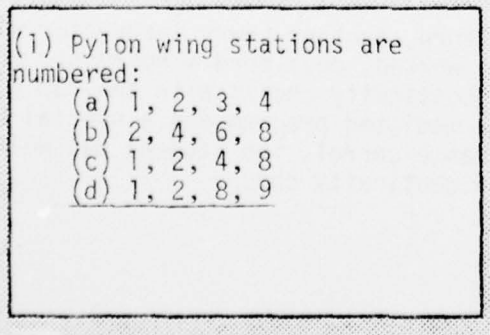


22

The student is given about 30 seconds for each test frame.



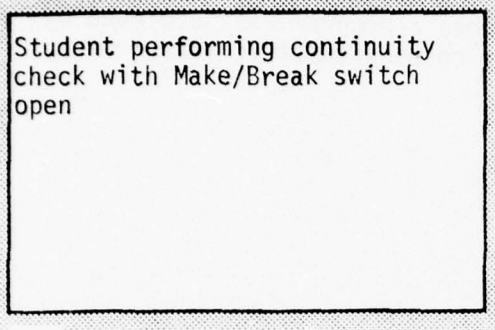
23



24

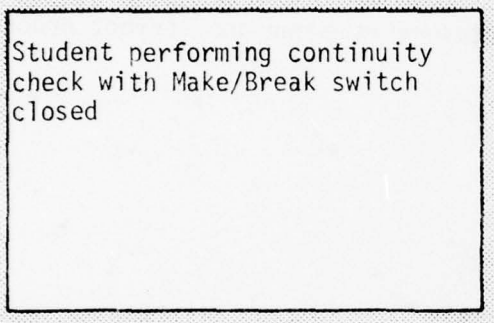
Then after he has completed the test he will then check the answers.

(8) The carrying hooks... (a) By gas pressure (b) By hydraulic pressure (c) By electrical solenoid (d) Pneumatically	
25	
(11) During ground operations .. (a) Turning off hydraulic pressure (b) Installing a lockout bolt (c) Installing a ground safety pin (d) Reversing breech sleeves	The sound track would say the answer is (c). During ground operations the accidental release of loaded stores is prevented by installing a ground safety pin.
26	
Instructor aiding student on MHU 83/E Program at standard carrel	Following each self test the student is invited to ask the instructor to explain any confusion which he may still have.
27	
PSM 6 Multi-meter with red lead being inserted	Mr. Raiford, another Lowry instructor with whom we have worked, developed a refresher module for making continuity checks with the PSM 6 Multi-meter. Using a mediated program and a special board in a performance carrel, the student was walked through a PSM 6 continuity check.
28	



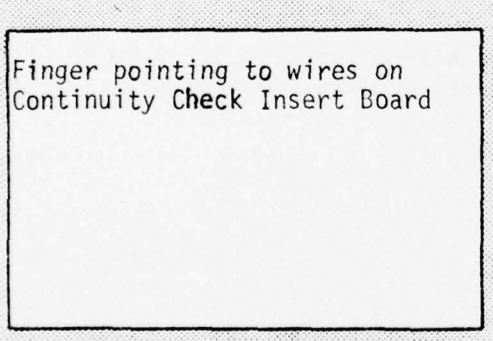
29

The student was shown an open circuit check which he then performed.



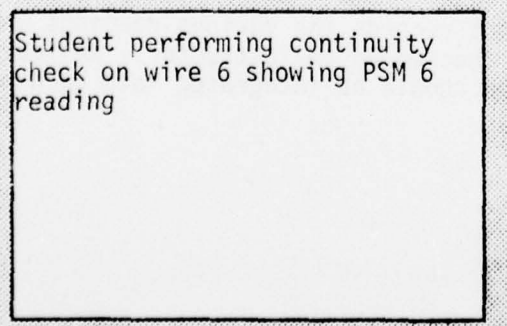
30

Then he was shown a closed circuit check which he performed.



31

After both open circuit and closed circuit examples were shown using wires, the



32

student then performed additional continuity checks. The results of proper checks were shown him. Successful completion of the program required the self-evaluation designed with the program.

DENSITOMETRIC MEASUREMENT

33

In another area a program was developed by Ron Filinger on densitometric measurement.

26. 21 STEP WEDGE - 11/11

Final frame in Densitometric Measurement Program. The student will measure 11 out of 11 steps on a 21 step wedge.

34

After the student is shown the correct procedure

Sensitometric Strip No. 0095

35

he is given a sensitometric strip on which he performs densitometric measurements.

Sensitometric Strip laying on Densitometric Measurement Answer Sheet

36

The student records the various readings which are then checked by an instructor. We feel that evaluation should be integrated into each module.

CARREL INSTRUCTION
QUALITY CONTROL
MODULE VALIDATION

Student working at standard carrel

37

Learning module validation will become increasingly objective. Modules will be systematically evaluated to insure optimal learning. The area of module evaluation will probably exercise the innovativeness and creativeness of military testers to the greatest extent.

Instructor focusing 1/2" video camera. MAU-12
Bomb Rack Program cues written on chalk board in background

38

In module development and validation we have been experimentally placing technology in the hands of experienced instructors. I will show you two of the techniques we have examined. The first technique is dynamic storyboarding using 1/2 inch video tape.

SHOOT

Instructor shooting Sony Rover 1/2" video camera in developing MAU 12 bomb rack module

39

An instructor shoots the desired performance of the job being trained. After review and feedback from other instructors, and possible revisions,

SHOW

Student viewing 12" video monitor in standard carrel over MAU 12 bomb rack module

40

the video sequence is shown to a small number of students. These students are

EVALUATE

Instructor evaluating student performance on MAU 12 bomb rack module

41

then taken to the equipment and their performance is evaluated. Any difficulties or confusion on the part of the student would necessitate some revision of the storyboard. When the module is informally validated, the media format for student learning is selected. We do not recommend the use of 1/2 video tape as a primary teaching media but normally would convert this dynamic storyboard to a prenarrated slide or Super 8 presentation. Revision or validation is completed before any production costs are incurred. When the instructor learns to develop modules which teach visually, another storyboard technique can be used.

SHOOT

Instructor developing MJ1 module is shooting with Insta-matic camera

42

Using an insta-matic or 35mm camera he can shoot selected critical views.

SEQUENCE

Instructor sequencing slides on multiplex slide sorter

43

Then when he receives the developed slides from the photo lab he selects and sequences a set of slides portraying what is to be taught.

SHOW

Instructor projecting slide from MJ 1 Module for students

44

He will then present the slide sequence to students - developing also the narrative associated with each visual.

SYNTHETIC JOB PERFORMANCE

Student operating switches in LT38 front cockpit

49

Evaluation of synthetic job performance can take many forms,

Vulcan Weapons System
CATB Program
Ft Carson, Colorado

Video camera setup which records panel during Vulcan Weapons System operation

50

For example at Ft Carson, video tape is used to evaluate the performance on the Vulcan Weapons System.

QUALITY CONTROL

Student performing arresting hook uplock assembly check on LT41

51

Quality control can be built into many modules. The instructor is able to make a performance check on each member of the team of four students who are learning the A/C preparation procedure for loading an AIM 7D guided missile in this quadraphonic sound module.

RELIABILITY

Student making continuity check in performance carrel

52

It seems quite important that the reliability,

EVALUATE

Instructor evaluating student performance on MJ 1 Module

45

Students can then be evaluated and the storyboar validated.

CARREL INSTRUCTION

QUALITY CONTROL

Student working at standard carrel

46

The third area of increased emphasis in evaluation related to carrel instruction is quality control. Quality control of criterion performances will be increasingly emphasized. Student sampling and extensive student performance tests can insure the objectivity and validity of student performance checks.

SYNTHETIC JOB ENVIRONMENT

Student sitting in front cockpit of LT38, a photo mockup with some live switches of F4C aircraft

47

Synthetic job environments can be readily created for comparative purposes. The use of synthetic performance tests, such as

COCKPIT PROCEDURES

Student at dummy rear cockpit panel, LT38

48

this cockpit procedures check, should increase because of their potential economy.

RELIABILITY
VALIDITY

Student making continuity check in performance carrel

53

validity and

RELIABILITY
VALIDITY
OBJECTIVITY

Student making continuity check in performance carrel

54

objectivity of all tests be established through student performance testing. An example of what I mean is now being conducted at Lowry AFB.

F4 Nose Gun ammunition handling system.
Pointing to link

55

We are asking the question of whether a picture test can substitute for a performance test in certain kinds of tasks. Using the same modules of instruction we are comparing two ways of testing.

INDIVIDUAL TEST
-PERFORMANCE-

Instructor testing student on MJ 1 parts location

56

Following exposure to the learning module Sgt Osborne takes the student to the actual equipment and asks the student to show him the selected items. The other way of testing permits group testing using pencil and paper.

GROUP TEST
-PENCIL PAPER-

Instructor testing students on MJ 1 parts identification

57

The student is shown a picture of the item and asked to select the correct name. These ways of testing are being assessed as to whether they can be considered equivalent.

CARREL INSTRUCTION

Student working at standard carrel

58

The use of carrel instruction appears as an opportunity for military testing

CARREL INSTRUCTION

QUALITY CONTROL

Student working at standard carrel

59

to provide better quality control of student performance,

CARREL INSTRUCTION

QUALITY CONTROL

MODULE VALIDATION

Student working at standard carrel

60

to systematically validate learning modules

CARREL INSTRUCTION

QUALITY CONTROL

MODULE VALIDATION

LEARNING EFFECTIVENESS

Student working at standard carrel

61

AFHRL
TECHNICAL TRAINING DIVISION
MEDIA DESIGN & EVALUATION
BRANCH

62

and to increase learning effectiveness.

Thank you

PROFICIENCY MEASUREMENT IN FLIGHT SIMULATORS

Edwin Cohen, Ph.D.
Simulation Products Division, The Singer Company
Binghamton, New York 13902

An increasing proportion of the work of military personnel involves participation in man-machine systems, such as aircraft, tanks, ships, missiles, and command and control systems, that are amenable to real-time, man-in-the-loop simulation. In such simulation, illustrated in Figure 1, the man is located in a replica of his normal work station, and provided with the same stimuli -- visual, aural, proprioceptive -- that comprise inputs to him during combat or other operations. He reacts to these stimuli, generating suitable outputs or responses -- control movements, switch actuations, verbal communications, log entries --; the simulator, which generally incorporates a digital computer, receives and acts upon these responses to modify the stimuli on a real-time basis. Some simulations are designed for crews, teams, or units, rather than single individuals. For some command and control or missile control simulations, the operational work station is used, with synthetic rather than real-world inputs being supplied. This is sometimes referred to as stimulation of the display system at the work station.

Positions amenable to simulation, although representing a small minority of armed forces personnel, contribute disproportionately to the effectiveness of each of the armed services. Because of the crucial nature of these positions, substantial efforts are invested in the selection and training of men to fill them, and in the assessment and verification of their combat readiness. Simulation plays a large and increasing role in training. For aircraft pilots, simulator training is far less costly than training in the aircraft; for decision makers in command and control systems, it allows practice of combat situations not otherwise available in peacetime. Simulators are also used for such civilian positions as airline flight crews, nuclear power plant operators, air traffic controllers, and automobile drivers.

Simulators have been a locus of measurement for a variety of purposes, indicated by the following sampling, which is meant to be illustrative rather than comprehensive:

1. Selection. AFHRL's Personnel Research Division has a program under way to determine the utility of the Singer GAT-1 simulator for selection of pilots for undergraduate pilot training.

2. Training Feedback. All current training simulators have provisions for furnishing the instructor with data on trainee performance. Some, such as the ASUPT, a T-37 simulator to be used by HRL at Williams AFB for training research, and the 2B24, a simulator of the UH-1 helicopter used for transition and instrument training at Ft. Rucker, furnish such data directly to the trainee in the cockpit.

3. Determining the Effect of Environmental Variables on Human Performance. Examples here include the effect of spacecraft-like confinement upon behavior, productivity with various work-rest cycles, and the effect of drugs upon driving and piloting behavior.

4. Optimizing and Validating the Design of Operational Systems. In the aircraft field, simulators are extensively used to check vehicle handling qualities and instrumentation.

5. Development of Procedures and Tactics. These range from communication procedures in CICs to air-to-air combat tactics.

Table 1 shows some of the advantages of simulation for various kinds of military systems; they can largely be summarized in terms of:

1. The breadth of situations that can be practiced,

2. The precise control of conditions of practice,

3. The easy availability of a broad variety of reliable, valid, and face-valid measures of man-machine system performance, and

4. The reasonableness of the costs associated with 1-3 above.

With all these advantages, why aren't simulators used even more as loci of psychometric effort? Much of the remainder of this paper will address that general question, focusing on flight simulators -- the kind with which I am most familiar -- although most of my remarks will be applicable to other kinds of simulators as well. As I see it, there are six factors making measurements of performance on simulators different from those to which psychometricians are accustomed, and hence inhibiting their use of simulators as loci of psychometric effort:

1. Historical Primacy of Paper-and-Pencil Tests. Most early testing, and certainly almost all successful early testing, used paper-and-pencil instruments; excursions into areas such as reaction time did not prove productive. Paper-and-pencil testing is cheap, and the domains it covers are

AD-A040 773 AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TEX F/G 5/9
PROCEEDINGS, ANNUAL CONFERENCE OF THE MILITARY TESTING ASSOCIAT--ETC(U)
NOV 73

UNCLASSIFIED NL

3 OF 9
ADA
040773



generally those in which the tester is comfortably superior to the testee. Thus the basic methodology of personnel testing and measurement has developed around paper-and-pencil instruments to the point where no one thinks it incongruous that "Tests and Measurements" textbooks devote upwards of 80% of their space to paper-and-pencil tests.

2. Multiplicity of Measures. When a psychometrician finally decides to gather human performance data on a simulator, he suffers an embarassment of riches -- scores or even hundreds of measures are available, all possessing face validity. Figure 2 shows data obtained during the landing of a 707 simulator; these items represent a selection from a much larger number of measures available.

3. Lack of Parallel Items. In a landing, there is just one indicated airspeed at touchdown; to get a parallel item, another landing must be conducted. This lack of parallel items within a given maneuver contrasts with the easy availability of parallel items for most paper-and-pencil test domains.

4. Difficulty in Combining Item Scores. With paper-and-pencil tests, the number of correctly answered items usually serves as an overall or summary measure; occasionally, differential weighting is employed. Only rarely is as complex an algorithm as successive hurdles used to derive a summary measure from item scores.

Much more complex algorithms are required to characterize proficiency in tasks such as landing a 707. For that landing task, items such as indicated airspeed, sink rate, distance from runway threshold, distance from runway centerline, and aircraft attitude (pitch, heading, and roll) -- all measured at touchdown -- cannot be evaluated independent of each other, as paper-and-pencil test items are generally evaluated, but must be combined, using a rather complex, yet to be developed algorithm.

5. Inapplicability of Conventional Item Analysis Techniques. Item analysis is most widely used for multiple-choice items. Each multiple choice item generally represents a separate behavior sample, and the alternatives are discrete. When, as they often do, measurements on a simulator cover various aspects of a single behavior sample, rather than separate behavior samples, and use continuous scales, rather than discrete alternatives, conventional item analysis is inapplicable.

6. Unavailability for Measurement of Behaviors Not Directly Reflected in Machine Inputs. In the simulator, the products of human behavior, as they affect the machine, are amenable to measurement, but the processes by which these

inputs are made are not. For example, in a flight simulator, we would normally have no way of knowing whether the pilot used an efficient visual scan pattern in checking instrument readings. The same objective performance could be achieved by two pilots with very different instrument scan patterns -- one using an efficient pattern and thus minimizing his work load, the other compensating for a less efficient pattern by working harder. When dealing with a simulator that enables measurement of such a large number of aspects of vehicle or system performance by merely specifying the parameters to be measured, it is all too easy to forget about those aspects of human performance not directly reflected in machine inputs.

By shifting our emphasis from paper-and-pencil testing to the kind of performance tests for which the simulator is such an appropriate vehicle, we would go far toward meeting the objectives of critics of our current testing posture, such as McClelland, who tells us to test for competence, rather than intelligence, and that "the best testing is criterion sampling" (McClelland, 1973, p. 7) and O'Leary, who advocates the "The advantages of job simulation approaches to selection over techniques lacking content validity" (O'Leary, 1973, p. 149).

In conclusion, we have a long way to go, in exploiting simulators as measurement tools, before their one basic weakness -- inability to measure behavior not directly reflected in machine inputs -- becomes a truly limiting factor. By refusing to be shackled to psychometric practices appropriate only to paper-and-pencil tests, we can exploit this powerful tool to measure a wide range of individual and team behavior, obtaining relevant, reliable, valid, and socially useful data at modest cost.

REFERENCES

1. McClelland, D. C. "Testing for Competence Rather Than for 'Intelligence'," American Psychologist, 1973, 28, 1-14.

2. O'Leary, L. R. "Fair Employment, Sound Psychometric Practice, and Reality. A Dilemma and a Partial Solution." American Psychologist, 1973, 28, 147-150.

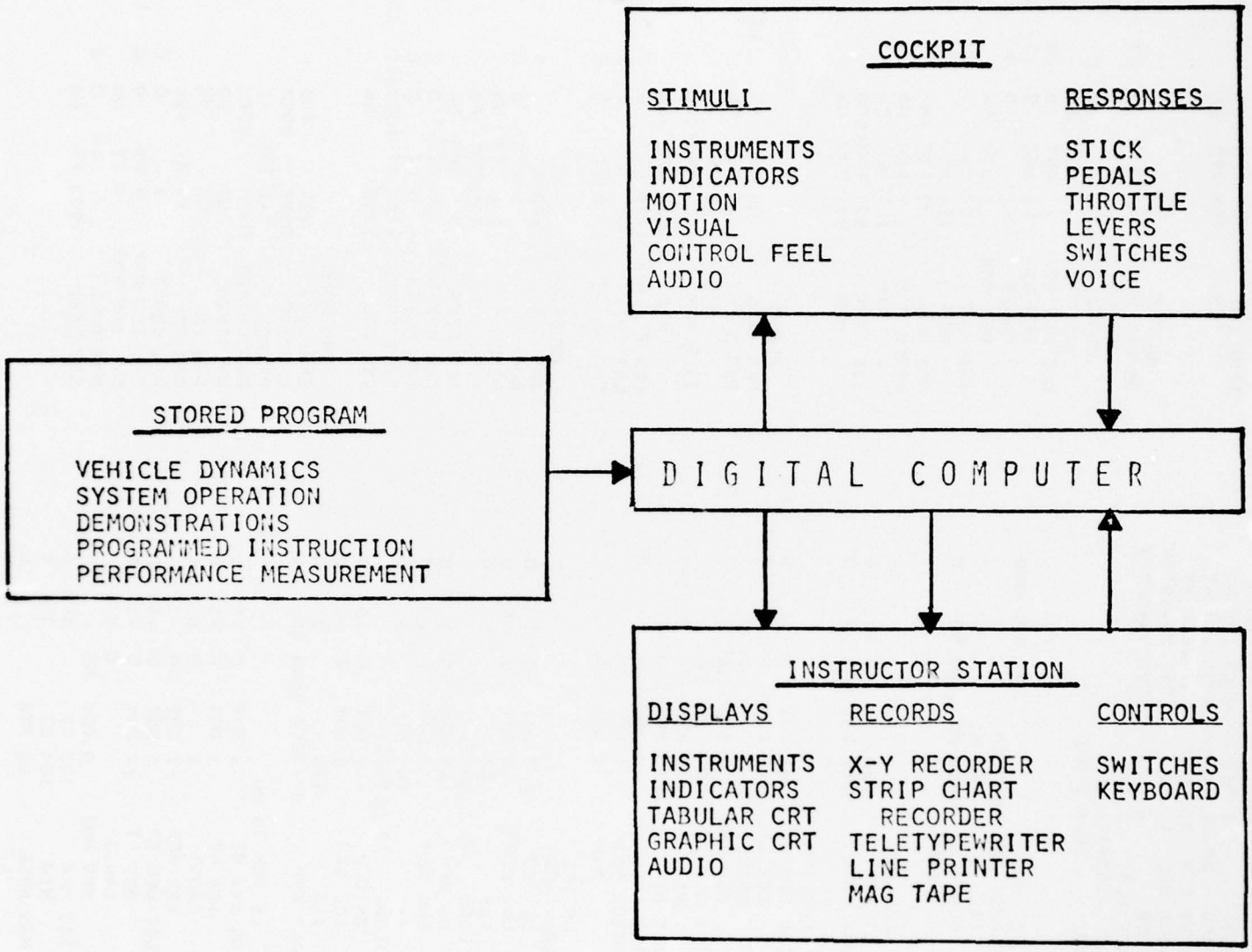


FIGURE 1. FLIGHT SIMULATOR BLOCK DIAGRAM

ILS LANDING MCI RWY 36

INITIAL COND		REF DATA	
OAT	6.1 DEG C	VREF	128 KT
BARO	29.91 IN HG	VMCA	117 KT
WIND DIR	36 DEG	ALT	2602 FT
WIND VEL	11 KT	LAT	39.1230N
TURB	0	LONG	94.8288W
RWY ICE	0 IN		
GW	206364 LB		
CG	30.4 % MAC		

DECEL		STD	TOL	DEV
FLAPS 25 DEGREES				
IAS	140 KT		5	3
ALT	2582 FT	2600	50	0
ALT DEV	0			
LOC INT				
ALT DEV	50	0	50	0
MAX ROLL	L21 DEG	<30	0	0
LOC TRK				
LOC DEV RT	.1 DOT	0	1	0
DEV LT	.1 DOT	0	1	0
CROSS		0	0	
DEV RMS	0 DOT	NA		
IAS MAX	147 KT	NA		
MIN	140 KT		0	-3
ALT MIN	2620 FT	2500	-50	0
MAX ROLL	L2 DEG	<20		0
PITCH MAX	5 DEG	NA		
MIN	2 DEG	NA		
GS INT				
GEAR DOWN				
GS DEV	-1.7 DOT	-2	1	0
LOC DEV	0 DOT	0	.5	0
IAS	142 KT		5	0
FLAPS 40 DEGREES				
GS DEV	-1.2 DOT	-9	1	0
LOC DEV	0 DOT	0	.5	0
IAS	140 KT		5	0
FLAPS 50 DEGREES				
GS DEV	-.1 DOT	0	.3	0
LOC DEV	0 DOT	0	.5	0
IAS	135 KT		5	0
ALT LOM	2497 FT	2500	50	0
GS TRK TO MM				
GS DEV HI	.1 DOT	0	.3	0
DEV LO	.1 DOT	0	.3	0
CROSS	1	0	0	1
DEV RMS	.1 DOT	NA		
LOC DEV RT	0 DOT	0	.5	0
DEV LT	.2 DOT	0	.5	0
CROSS		0	0	
DEV RMS	0 DOT	NA		
IAS MAX	133 KT		5	0
MIN	124 KT		5	0
MAX ROLL	L2 DEG			0
PITCH MAX	2.5 DEG	NA		
MIN	-.7 DEG	NA		
SINK MAX	725 FPM	<999		0
MIN	469 FPM	NA		
GS TRK				
GS DEV H	.2 DOT	0	.3	0
DEV LO	.1 DOT	0		0
CROSS	1	0	0	1
DEV RMS	.2 DOT	NA		
LOC DEV RT	.1 DOT	0	.3	0
DEV LT	.1 DOT	0	.3	0
CROSS	1	0	0	1
DEV RMS	.1 DOT	NA		
IAS MAX	129 KT		5	-4
MIN	124 KT		5	0
MAX ROLL	R3 DEG	<5		0
PITCH MAX	2.9 DEG	NA		
MIN	1.1 DEG	NA		
SINK MAX	573 FPM	<800		0
MIN	371 FPM	NA		
FLARE				
MAX ROLL	R5 DEG	<3		L2.5
PITCH MAX	2.7 DEG	NA		
MIN	.2 DEG	NA		
SINK MAX	635 FPM	<800		0
MIN	404 FPM	NA		
DEV RWY HDG	R7.3 DEG	0	2	5.3
LAT DRIFT	L11 FPS	0	0	11
C/L ALIGN	0 FT	0	0	0
TOUCHDOWN				
IAS	122 KT		5	-1
ROLL	L DEG	0	2	0
PITCH	1.2 DEG	NA		
SINK	505 FPM	NA		
ACT SINK	509 FPM	NA		
DEV RWYHDG	R6 DEG	0	2	-4
C/L ALIGN	0 FT	0	0	0
DIST THRESH	1727 FT	1000	500	225
THR POS	IDLE	NA		
GRD ROLL				
SPOILERS	YES	YES		
BRAKING	DIFF	NORM		
RESPONSE	2.2 SEC	<5		0
NW CONTACT	NORM	NORM		
NW STEER	NO	NO		
REV FN	NO	NO		
MAX N1	76 %	<94		0
C/L CROSS	1	0	0	1
C/L DEV MAX	8 FT	0	0	L8
C/L DEV RMS	0 FT	NA		
MAX ROLL	DEG	0	3	
RWY USED	1900 FT	NA		

The teletypewriter printout actually took the form of a single long column; it is formatted in two columns to save space.

Figure 2 DATA OBTAINED DURING LANDING OF A 707 SIMULATOR

TABLE 1: ADVANTAGES OF SIMULATORS OF VARIOUS KINDS OF MILITARY SYSTEMS

ADVANTAGES	TYPE OF SIMULATOR						
	FLIGHT	AIRCRAFT WEAPON SYSTEM	TANK	SUB DIVING	AIR DEFENSE SYSTEM	ICMB LAUNCH CONTROL	COMBINED ARMS TACTICAL TRAINER
COST SAVINGS	X	X	X	X	X	X	X
REDUCTION OR ELIMINATION OF DELAYS INHERENT IN USE OF OPERATIONAL SYSTEM	X	X	X	X	X	X	X
INDEPENDENCE OF SEQUENCING REQUIRED BY OPERATIONAL SYSTEM	X	X	X	X	X	X	X
ABILITY TO SPECIFY ENVIRONMENT: WIND, WEATHER, SEA STATE, VISIBILITY, ETC.	X	X	X	X	X	X	X
FREEDOM FROM CANCELLATION DUE TO WEATHER, EQUIPMENT MALFUNCTION	X	X	X	X	X	X	X
SAFETY	X	X		X	X	X	X
SIMULATED MALFUNCTIONS	X	X	X	X	X	X	X
ANY GEOGRAPHIC AREA	X	X	X	X	X	X	X
REALISTIC ENEMY		X	X	X	X	X	X
INTELLIGENCE SECURITY		X	X		X	X	X
NO RESTRICTIONS ON WEAPON FIRING		X	X			X	X
OBJECTIVE MEASURES OF MAN-MCAHINE SYSTEM PERFORMANCE	X	X	X	X	X	X	X
NO RANGE OR AIRSPACE LIMITATIONS	X	X	X		X	X	X
NO ENVIRONMENT IMPACT	X	X	X		X	X	X
SIMULATOR AVAILABLE BEFORE SYSTEM OPERATIONAL	X	X	X	X	X	X	X

THE PERSONNEL AND TRAINING EVALUATION PROGRAM: A Working Program for Improving the Efficiency and Effectiveness of Fleet Ballistic Missile Weapons System Training

Part I Scope and Achievement

by
Lcdr Clarence L. Walker USN

Central Test Site for the Personnel
and Training Evaluation Program

INTRODUCTION

In April of 1969 the Chief of Naval Operations issued an instruction which brought into being the Fleet Ballistic Missile Weapons System Training Program. This Program was aimed at improving the effectiveness and efficiency of training in the areas of the missile, missile launcher, missile fire control and navigation on FBM submarines and at their support activities. Specific gains looked for were the shortening of the replacement pipeline, provision for advanced, vice refresher training at the ship's homeports and standardization of the advanced training available. A built-in part of the FBM Weapons System Training Program was a monitor called the Personnel and Training Evaluation Program (PTEP). The program:

> "provides the organization, procedures, and responsibilities required to accomplish the qualitative assessment of personnel knowledge and skill levels for officer and enlisted personnel during replacement training and while assigned to SSBNs, FBM Tenders, and FBM training facilities. The PTEP also provides for evaluation of training facilities, hardware, documentation, and courses of instruction for use as the basis for implementing improvements in training and in all elements of the FBM Weapons System Training Program."

Various aspects of PTEP have been presented at the last two meetings of this association. For these presentations PTEP was in the speculative stage. This paper addresses the early working stages. To understand PTEP, however, it is necessary to look at the climate in which it operates.

PTEP is presently limited to the Poseidon Missile community, which consists of:

1. A replacement training site - Naval Guided Missiles School
   Dam Neck, Virginia
   Dam Neck, Virginia
2. Two off-crew training sites - Naval Submarine School,
   Groton, CT.
   FBM Submarine Training Center,
   Charleston, SC.
3. Two submarine tenders
4. Twenty-six FBM Submarines each with 2 crews

The submarine crews comprise the majority of the personnel and absorb the vast majority of the effort.

FBM submarines utilize a two-crew concept: One crew mans and operates the submarine while the other crew is available at an off-crew site for training. Planning revolves around a six-month patrol cycle. This consists of one patrol/refit period and one training/leave period. Each man remains aboard for approximately six patrol cycles or three years; he is then rotated to shore duty, frequently at a training site; following this he returns to sea again. All personnel selected for FBM Weapons and Navigation Training are of above average intelligence. The submarine personnel have an initial six year obligation, with the exception of the Torpedomen responsible for the launch tubes. First-term reenlistments run in excess of 40%.

These then are the personnel involved in the FBM Weapons System Training Program. The training program itself is not unique in its academic concepts, but it is innovative in the area of program management. As with any viable training system, the FBM Weapons System Training Program first provides training objectives. It determines where, when and by whom the learning is to take place; it provides the learning opportunity, and then assesses the quality of the end product.

PERSONNEL PERFORMANCE PROFILES AND TRAINING PATH SYSTEM (PPP)

The Personnel Performance Profile (PPP) is the basic building block of the system. The profiles are for the most part developed by hardware contractors and state in a standardized format the skills and knowledge necessary to operate and maintain a system or equipment.

A system called The Training Path System assigns to each FBM Navy Enlisted Classification (NEC) the profile items and levels of achievement appropriate. This same system, by use of profile items and levels of achievement, assigns material to replacement, advanced and on-board training. Curricula are prepared using the profiles and Training Path System. Informal training materials in support of the profiles have been developed, and watch qualifications requirements have been keyed to the profiles and levels of achievement. The FBM Weapons System Training Program provides a system, whereby, the required knowledge for an individual or the specified content of a course can be conveniently identified by listing profile items and achievement levels.

PTEP ADMINISTRATION

The Personnel and Training Evaluation Program is administered by 35 military personnel and a contractor civilian staff of about 10 people. One officer and an enlisted staff of experienced technicians are stationed at each training site to support the testing and evaluation tasks. The bulk of the enlisted staff is at the Central Test Site in Dam Neck, Virginia. The program has a broad charter of evaluation. Evaluation requires facts. PTEP gathers its facts through examinations, collection of personal training information, ship and equipment operational history, and review of curricula with their supporting material. Examinations are the primary source of data.

Two separate types of examinations are administered by PTEP. One measures the knowledge of an individual against knowledge required of him by the Training Path System. This is called a System Achievement Test, or SAT. The other type of examination measures the performance of a course against its Training Path System requirements. This is called a Course Achievement Test or CAT. In both cases examination content is based on PPP items and achievement levels prescribed by the Training Path System.

Course Achievement Tests

Course Achievement Tests are designed to measure both the performance of an individual in a course and the performance of the course against its prescribed content. The instructor does not see the test until it is time to administer it and so has no opportunity to "teach" the test. Standardization of training between sites can also be measured

with a Course Achievement Test. Few CAT's have been produced to date primarily because the standardized courses they support have not been in place. No meaningful analysis has been possible with the results.

System Achievement Tests

The greatest field of endeavor, and heart of the Personnel and Training Evaluation System from the testing point of view, is System Achievement Testing. The tests were implemented incrementally by NEC over a period of 9 months, starting in August 1971. To date there have been more than 4,000 administered. Each of the several tests in use is designed to cover the full scope of the knowledge required for a specific Navy Enlisted Classification. Content is determined using the Personnel Proficiency Profiles, and the Training Path System. The tests consist of two parts, a knowledge part consisting of from 180 to 360 multiple choice questions and a skill part whih uses a paper work fault isolation technique which has been developed by the Data-Design Laboratories. Depending on the NEC involved the tests take 3 or 4 three hour testing sessions. Each of the testing sites is connected by a data link with a central computer complex. The tests are scored, training recommended, and exam question information stored for future analysis by the computer. Test results are returned to the testing site by teletype. From the outset these SAT's have proved to be a reliable indicator of individual knowledge. The primary measure of test validity has been a comparison of test results with the subjective analysis of the individual concerned by their immediate supervisors. No serious discrepancies have been noted. The System Achievement Test, however, is much more than a

vehicle to confirm opinions readily derived by asking supervisors. This can be seen from the nature and construction of the tests. (Figure 1 Sample Test Result)

Each test is structured to provide information on the sub areas which make up an NEC and to recommend specific remedial action where indicated. Analysis of individual areas provides information on the effectiveness of training in each of the areas.

Personnel Data System

The Personnel Data System is the part of the Personnel and Training Evaluation Program which changes a useful testing scheme into a valuable evaluation tool. This system picks up a man as he leaves replacement training. It records each man's training and duty station history and provides the information required for measuring the effectiveness of training at various locations and on personnel with varying training backgrounds. It also provides the statistical base against which achievement or lack of it can be measured. At present there are over 3,000 records in the Personnel Data System. (Figure 2 Personnel Data Sheet)

One can thus see that PTEP provides the capability for collecting information on the training and personnel performance for which the FBM Weapons System Training Program is responsible. It also provides a broad source of background material against which to evaluate this information.

PTEP ACCEPTANCE

When first introduced to the fleet, System Achievement Testing was not met with overwhelming enthusiasm. Guarded skeptism is about as close as anyone came to acceptance. The rigorous schedule of examinations already participated in by FBM submarine crews in the areas of nuclear propulsion and nuclear weapons safety contributed to this unenthusiastic reception.

Gaining acceptance for the System Achievement Test therefore was not a simple job. A strong public relations effort was launched which started with the indoctrination on the FBM Weapons System Training Program of everyone taking the exam as well as the administrative personnel responsible for using the results. To further aid both in acceptance of testing and the use of test results, a policy was adopted whereby only the command examined and not its superiors in the chain of command was supplied with the test results. This gave each commanding officer the capability of evaluating examination results in the light of his own knowledge of his ship's training needs and to make judicious use of the results. The practice removed from both the commanding officer and the PTEP organization the requirement to defend results from an untried test instrument.

Perhaps the most important step toward acceptance was the procedure whereby each set of test results was returned to the command by an officer assigned PTEP duties. The results were discussed with the commanding officer and department heads in the light of individual and overall crew performance. Carefully kept records allowed these PTEP officers to review trends and to point out consistent low and

high performers as well as anamolous results. The personal and individual interest shown by the PTEP organizations at the training sites effectively sold the testing system to the ships.

Training recommendations were at first reluctantly used by the submarines. An early sampling showed that courses recommended by PTEP were used at about the same rate as randomly chosen courses. A graphical analysis of successive test results presented a convincing picture that use of training courses improved performance in the area where the course applied; and that the most gain for training time expended was realized when the time was applied to weak areas. This seemingly self evident conclusion, convincingly demonstrated to management personnel, has much increased the utilization of PTEP course recommendations. The use of self-study material recommended by PTEP was another area that gained slow acceptance. As with courses, however, PTEP has been able to demonstrate that men who use self-study material show improvement in their overall performance; consequently the use of this training material has increased considerably.

ANALYSIS AND EVALUATION

PTEP's business is much more than improving training utilization. Since January of 1973 sixteen evaluations have been completed on such diverse topics as: curriculum management materials, fire control software training, training of submarine tender personnel, and the restructuring of a sonar training course. Evaluation of test results of individuals and crews have shown that personal and leadership problems are reflected in performance trends and anomalies;

these results could give early indication of personnel problems. It must be pointed out that to commence and continue a program such as PTEP rigorous and continuing internal evaluation of material, procedures and current objectives are necessary. Much attention has been focused on this facet of PTEP. Due to the nature of PTEP, external suggestions for improvement are not tardy in coming either.

OPERATIONAL CRITERIA

The Personnel and Training Evaluation Program described thus far is one which closes a neat academic loop--do poorly on our test, take our instruction and you will do better on our next test.

The education provided, however, is not an end in itself, but a means to an end. That end is the effective operation of the Fleet Ballistic Missile Weapons System. Correlation of training with actual operations is of paramount importance in measuring the effectiveness of training. Since PTEP can accurately reflect training, comparing PTEP results with operational information should link the training and operating worlds. FBM Submarine patrols and equipment operation are meticulously documented. Equipment failures are analyzed, as are operational procedures and repair procedures by sources outside PTEP. From this wealth of data, information should be available to approach criterion testing backed by operational statistics. The problem, in this continuing effort, is sorting out the multitude of difficult-to-relate variables in order to correlate operational information with testing information. To date, PTEP has found correlation in test and operational trends. Spearman Rank Difference correlations have been positive and have run as

high as +.83. By July of 1973 correlation of test results with operational information was sufficiently strong to warrant the release of test results to the Submarine Group Commanders to assist them in evaluating the personnel and training needs of their subordinate commands. PTEP can indicate crews with an increased probability of difficulty in handling material problems arising at sea. It is in no way able to predict casualties.

CURRENT STATUS

What has been the net impact of PTEP on the ships supported by the FBM Weapons System Training Program? It has provided a positive means of identifying the best means of employing training time. It has indicated where serious training deficiencies exist, particularly in supervisory personnel. It has provided a tool for identifying to higher commands personnel and training deficiencies beyond the immediate control of the submarine Commanding Officer.

For the training command, PTEP has standardized definitions of objective and insisted upon compliance with these objectives. For management commands, PTEP has demonstrated the effectiveness and ineffectiveness of various management tools built into the FBM Weapons System Training Program. It has provided a quickly accessible pool of information for analyzing crew, rate, or NEC problems. PTEP has the capability and has made beginnings towards measuring the product of the present FBM Weapons System replacement training program against the product of the previous program. Both the Naval Examining Center and the Bureau of Naval Personnel have been provided with PTEP products.

PTEP has made a constructive contribution to the development and implementation of the FBM Weapons System Training Program. Its importance in maintaining a responsive training system continues to grow as the Training Program grows. With the addition of the automated access to information files, which is expected before the end of the year, the possibilities for analyzing many facets of the complex personnel and system inter-relationships of military training become awesome.

08/16/73 GROUP SAT REPORT SSBN 695 BLUE CR

NEC 3306 FIRE CTRL TECH SSBN TEST NUMBER F17304 TESTED 73/08/06

KNOWLEDGE AREAS-PART 01

A FBM WEAPON SYSTEM & EQUIPMENTS
B CONT CONSOLE & PWR SUBSYSTEMS
C PLATFORM POSITIONING EQUIPMENT
D MSL & GUID TESTING EQUIPMENT
E DIGITAL CONTROL COMPUTER
F MULTIPLEXED EQUIPMENT
G DCC - OPERATOR INTERFACE
H DCC TO MISSILE INTERFACE
I MAGNETIC DISC FILE
J OPTICAL ALIGNMENT GROUP
K MTRE MK 6 MOD 3

AVERAGES	KNOW	SKILL
1 FLEET	50	50
A E7/E8	57	52
B E6	58	54
C E5	51	50
D E3/E4	43	45
2 GROUP	49	53

SKILL AREAS-PART 02

L DIGITAL CONTROL COMPUTER
M TEST SUBSYSTEM
N DCC PERIPHERAL EQUIPMENT

NAME P/G	KNOWL SCORE	SKILL SCORE	AREA SCORES A	B	C	D	E	F	G	H	I	J	K	L	M	N
GROUP	51	49	50	56	50	55	55	52	47	47	54	51	48	45	52	49
PUCKETT, B.			652-94-4234													
E4	35	27	32	41	42	40	46	37	30	30	54	47	40	24	33	35
RUSHIN, J.			328-82-9213													
E5	51	50	57	56	47	55	55	47	47	47	42	51	52	54	50	47
SCOTT, L.			747-42-6469													
E6	68	68	62	67	62	71	62	69	64	64	63	55	52	56	71	66

Figure 1. Sample SAT Report

DATE 08/14/74 PREP PERSONNEL DATA SHEET

NAME	SSAN	RANK/ RATE,DOR 003	PRANK/ PRATE,DOR 004	SUBSPECIALTY/ PNEC,DATE 005	SUBSPECIALTY/ SNEC,DATE 006	TNEC,DATE 007
BROWN JOHN CAMERON	136-34-6438	ETN2 711101	ETN3 691009	3325 6911	0000	0000

TPC 009	POLARIS 010	POSEIDON 011	DUTY STATION,DROB 012	INST EXP 013	CONST/CONV 014	CIV ED 015	PRD 016	LOSS/ EAOS 017
N51	00 PATROLS	34 PATROLS	SSBN 627 BLUE CR 6911	00 MONTHS	16 MONTHS	13 YEARS	7405	740715

SSQUAL,REQL 018	WATCH 020/024	DUTY STATION HIST 030/034	BASIC BATTERY SCORES 035/047				OFF-DUTY COURSES,DOC 050/054	
7102	020	030 GMS DAM NECK 6809	035 AFQT 95		042 RADIO 67		050 PSYCHOLOGY 201	7107
	021	031 RTC GREAT LAKES 6807	036 GCT 73		043 ETST 73		051	
	022	032	037 ART 66		044 READ		052	
	023	033	038 MECH 65		045 MATH		053	
	024	034	039 CLER 41		046 PHYS		054	
			040 SPT 63		047 ELEC			
			041 SONAR 70					

NAVY SCHOOLS,DOC 055/064

	NUMBER	TITLE	C/S	DOC
055	E193089CNCPROD1		C2/35	73C6
056	E193028MSRADM		C1/04	7201
057	E193020SINSCMA	/SAT/		7111
058	E193018STASTAC		C1/C3	71C7
059	E193077INAVT/A		C8/18	70C9
060				
061				
062				
063				
064				

SSBN COMPLETIONS 065/084

	SSBN	VOL	PT REV	DOC
065	N16			7305
066	N17			7209
067				
068				
069				
070				
071				
072				
073				
074				
075				
076				
077				
078				
079				
080				
081				
082				
083				
084				

CAT RECORD 100/119

	TEST NAME	DATE	SCORE
100	946967301	730425	58
101			
102			
103			
104			
105			
106			
107			
108			
109			
110			
111			
112			
113			
114			
115			
116			
117			
118			
119			

SAT RECORD 120/129

	TEST NUMBER	DATE	KNOW SCOR	SKILL SCOR	DAY
120	NS173C5C3	73C7C3	66	55	02
121	NS1721201	730117	58	59	02
122	NS1720601	720710	52	56	02
123					
124					
125					
126					
127					
128					
129					

MULTIMEDIA SETS 130/139

	MMS	REV	CFG	DOC
130				
131				
132				
133				
134				
135				
136				
137				
138				
139				

Figure 2. Sample Personnel Data Sheet

THE PERSONNEL AND TRAINING EVALUATION PROGRAM: A Working Program for Improving the Efficiency and Effectiveness of Fleet Ballistic Missile Weapons System Training

Part II Program Development

by
Frank B. Braun

Data-Design Laboratories
Norfolk, Virginia

The purpose of this paper is to discuss some of the problems and decision points encountered in the development and implementation of the Personnel and Training Evaluation Program (PTEP). As in the case of any system development, not all the decisions made in the heat of battle were right or the most appropriate. Many of our solutions may be of interest, however, since we do have a working program for testing and evaluating military technical specialists.

PTEP is basically the product of system engineering. The personnel involved in the development of the program are engineers, instructors, and technical specialists; a systems analyst and a programmer applied automatic data processing (ADP) procedures to almost all aspects of the program.

Since "who" we were going to evaluate was pre-determined, i.e., the FBM weapons and navigation technicians, the first problem we had to tackle was what we were going to use to perform the evaluation. We started with an advantage, since a job task analysis already existed in the form of the Personnel Performance Profiles (PPP) and the Training Path System (TPS). The PPP and TPS provide the knowledge

and skill requirements for each technician in the program, and are used as the standards for all elements of the program.

We began our task by searching for the most appropriate types of test instruments to test the PPP requirements. Review of the many reference sources on the subject lead us to the decision that the four-alternative, multiple-choice question would be the most satisfactory for testing knowledge items. Since the questions would be acquired from many different sources, i.e., equipment manufacturers and various Navy school instructors, a specification was developed to insure test item standardization. Along with the question, the test item writer is required to provide various information to relate the question to the PPP/TPS standards (Figure 1). For the skill items contained in the PPP/TPS, it was determined that performance tests would be the most desirable test instruments, but were impractical for a large scale testing program. However, Decision Development System (DDS) exercises, reported on at the 1970 and 1972 MTA meetings, could be keyed to the PPP/TPS standards and appeared to be a satisfactory means of testing the technician's application of knowledge. We decided to utilize the DDS as the skill test instrument.

The next task was to develop tests from the PPP/TPS requirements which would measure each technician's knowledge and skill level. The problem that became apparent in this area was the wide range of requirements for each technician. In the case of the Fire Control Technician, for example, there are 28 primary equipment profile tables, 10 secondary equipment profile tables, and two system profile tables. Each profile contains approximately 35 knowledge and skill items with varying

ACTION	TABLE NO.	ITEM NO.	SEQUENCE NO.	FORM	SUB-ITEM NO.	EXAMINEE SYS.	EXAMINEE TYPE	TOS LEVEL	EQUIPMENT Mk-Mod	SEC. CLASS.	SPEC. HAND.	SUPP. CODE	ADMIN. APPLIC.
000	001	002	003	004	005	006	007	008	009	010	011	012	013
	512	1612	7104	0		C	T	T2	A/D	U		DN	MC

REFERENCE PUBLICATION NO.	PUBLICATION DESCRIPTION VOL.	PART	REV.	CHG.	PUBLICATION BREAKDOWN TYPE	NUMBER	DATE YY MM DD	FIGURE NUMBER	ANS.	TEXT LENGTH
014	015	016	017	018	019	020	021	022	023	024
0652-004-8	1			6	P	4-5	73 07 06		1	06

7309BA

A/D CONVERTER. WHERE IS THE PRE-OPERATIONAL CHECKOUT PROCEDURE FOR THE A/D CONVERTER DESCRIBED?

1. A/D CONVERTER TECH. MANUAL
2. FBM STD. MAINTENANCE PROCEDURES
3. FBM STD. NAVIGATION OPERATING PROCEDURES
4. NAVIGATION SUBSYSTEM MANUAL

Figure 1 - SAMPLE TEST ITEM

numbers of sub-items under each item (Figure 2). Obviously, we could not test each and every requirement within a reasonable length of time. Our solution to this was the System Achievement Test (SAT) which consists of both a knowledge and a skill part. The knowledge part of this test is generated using a weighted random sampling method. Each technician's PPP tables are assigned weighting factors based on their relative importance to the overall requirements. Each PPP item within a table is also assigned a weighting factor based on complexity and importance in relation to the other items in the table. When a test is generated, the computer uses the weighting factors to develop a sampling bank of all applicable questions. Then, a random generator selects from the sampling bank and generates the knowledge part of the test version. The skill part of each SAT is generated manually after a thorough review of the DDS exercises applicable to the PPP standards. Each test version undergoes a thorough engineering review by system experts and is administered to a pilot group for further refining prior to issuing as a new test.

Another type of test was developed to use in the training courses administered at the FBM Training Activities located at the Guided Missiles School, Dam Neck, Virginia, the Fleet Ballistic Missile Submarine Training Center, Charleston, South Carolina, and the Naval Submarine School, New London, Connecticut. This test, named the Course Achievement Test (CAT), presented a different set of problems. The PPP items covered by each curriculum are indicated in a chart which relates course objectives to specific PPP items (Figure 3). This chart, the Objective Assignment Chart (OAC), becomes the basis for designing the CAT for the training course. It was decided to make the CAT a criteria-referenced test since the standards could be defined

TRAINING PATH CHART FOR FIRE CONTROL TECHNICIAN (SSBN) POSEIDON,
FCS MK 88 (NEC FT-3306) (SHEET 2 OF 2) TPC-F1

TABLE INDEX

010 FBM Weapon System (System Level)
100 Fire Control Systems - Mk 88 Mods 0 and 1 and Mk 84 Mod 1 (Subsystem Level)
105 Alignment Subsystem - Mk 88 Mods 0 and 1 and Mk 84 Mod 1
106 Erection Subsystem - Mk 88 Mods 0 and 1 and Mk 84 Mod 1
107 Missile Motion Subsystem - Mk 88 Mods 0 and 1 and Mk 84 Mod 1
108 Digital Read-In Subsystem - Mk 88 Mods 0 and 1 and Mk 84 Mod 1
109 Master Clock and Timing Subsystem - Mk 88 Mods 0 and 1 and Mk 84 Mod 1
110 Command Slew Subsystem - Mk 88 Mods 0 and 1
111 Control and Display Subsystem - Mk 88 Mods 0 and 1 and Mk 84 Mod 1
112 Rate Compensation Electronics - Mk 88 Mod 1
113 Automatic Fault Registration Subsystem - Mk 88 Mods 0 and 1
114 Multiplexers - Mk 88 Mods 0 and 1 and Mk 84 Mod 1
115 Digital Control Computer - Mk 88 Mods 0 and 1 and Mk 84 Mod 1
116 Digital Control Computer to Equipment Configuration Switching and Patching - Mk 88 Mods 0 and 1
118 Servo Group Subsystem - Mk 88 Mods 0 and 1
120 Fuze Set Subsystem - Mk 88 Mods 0 and 1 and Mk 84 Mod 1
121 Parallel-To-Serial Converter - Mk 88 Mods 0 and 1 and Mk 84 Mod 1
122 Events Translator - Mk 88 Mods 0 and 1
124 Keyboard and Keyboard Display - Mk 88 Mods 0 and 1
126 Intercomputer Link - Mk 88 Mods 0 and 1
127 Magnetic Disk File - Mk 88 Mods 0 and 1
131 Computer Printer - Mk 88 Mods 0 and 1 and Mk 84 Mod 1
133 Computer Tape Reader Mk 1 Mods 0 and 2 - Mk 88 Mods 0 and 1, Mk 80 Mod 2, and Mk 84 Mod 1
135 Tape Reader Test Set
137 Power Distribution Subsystem Mk 88 Mods 0 and 1 and Mk 84 Mod 1
140 Test Subsystem - Mk 88 Mods 0 and 1
145 Test Instrumentation Adapter - Mk 88 Mods 0 and 1
146 Digital Transmitter Receiver - Mk 88 Mod 1
151 Temperature Monitoring Power Supply Mk 141 Mod 1 - Mk 88 Mods 0 and 1
155 Optical Alignment Group Mk 2 Mod 0 - Mk 88 Mods 0 and 1
189 Module Test Set Mk 412 Mods 0 and 1 - Mk 88 Mods 0 and 1 and Mk 84 Mod 1
200 Missile Systems (Subsystem Level)
204 Guidance Subsystems
205 Gimbal Assemblies Mk 3 Mod 0 and Mk 2 Mod 1 and Associated Electronics
206 Guidance Computer Mk 3 Mod 0
310 Missile Test and Readiness Equipment Mk 7 Mods 1, 2, and 3
315 Missile Test and Readiness Equipment Mk 6 Mods 1 and 3
400 Missile Launching Systems (Subsystem Level)
500 Navigation Subsystems (Subsystem Level)
508 Multispeed Repeater Mk 3 Mod 3
540 Frequency Time Standard AN/BSQ 2, -2A, and -2B
541 Optical Equipment
600 Weapon System Ship Support Subsystem (Subsystem Level)
610 Ventilation System
620 Electrical System

Figure 2 - SAMPLE PPP TABLE INDEX

TABLE	ITEM	TOS	VOL.	PART	SECT.	TOPIC	TOPIC OBJ.	TABLE	ITEM	TOS	VOL.	PART	SECT.	TOPIC	TOPIC OBJ.
635	1-1-1	T1	3		1	1	1								
	1-1-2	T2	3		1	1	2								
	1-1-3	T2	3		1	1	4								
	1-1-4	T2	3		1	1	3								
	1-2-1	T2	3		1	2	1								
	1-2-2	T1	3		1	2	2								
	1-3-1	T2	3		1	3	1								
	1-3-2	T2	3		1	3	2								
	1-4-1	T2	3		1	4	1								
	1-4-2	T2	3		1	4	2								

Figure 3 - SAMPLE OBJECTIVE ASSIGNMENT CHART

adequately to support this type of test. Our biggest problem to date has been in obtaining stabilized, approved curricula for which tests can be designed. These curricula are now becoming available, and the CAT program is finally gathering momentum.

Many problems arose in the actual administration of the tests. Three distinct environments existed which required different approaches to administering the tests. First, we had the submarine crew during its off-crew period. It was decided to use a dedicated Navy team at each site to administer the tests. This solved the potential problem of having civilian engineers give tests to military personnel. Every attempt was made to avoid the "Big Brother is watching you" image in order to obtain the best possible, positive attitude towards the tests by the examinees.

The second testing situation was with the submarine tender personnel. Since the tenders were located at remote sites, test packages were developed which could be administered by the tender personnel themselves. This meant providing clear, easily understandable instructions for each test section and directions for return of the packages upon completion of the test.

The third testing situation involved the CAT. This test is administered by the instructor at the scheduled time in the course. For the CAT, we developed a simple proctor guide and a punched, overlay type answer key. These materials, along with the test, are provided by the local Navy team to the instructor on the scheduled examination day. This procedure prevents the instructor from teaching the test.

The System Achievement Test, because of its objective to test the technicians' total system capabilities, is a relatively lengthy test. The SAT ranges from one 3-hour session to four such sessions, depending on the particular PPP/TPS requirements. Obviously with a test this long, we were concerned with the attitude of the personnel required to take it. Surprisingly, after the initial grumblings, the technicians appear to view the SAT as an integral part of their training program. Some technicians have even provided recommendations to improve various aspects of the tests.

The DDS exercises require the use of the same technical documentation employed by the technicians when troubleshooting similar casualties onboard ship. We decided early in the development of the program that close monitoring of the change status of the publications would be required to insure currency of the tests. It was also decided that the local Navy team at each site should maintain its own technical library to insure positive control of the change status of the publications used with the tests. This decision created extra work for the teams but reduced the cries of "this won't work" during the test sessions.

During the first year of testing, the raw results were mailed from the test site to the central site for scoring. This created a delay of 10 to 15 days from the completion of the test to the time the results were returned. The program managers in the Strategic Systems Project Office directed that this turnaround time be reduced to enable commands to schedule remedial training, if the test results indicated a significant deficiency, prior to departure for patrol. An optical scanning device and a teletype machine were used to solve this problem. The Navy team processes the examinees' answer sheets through the optical

scanning machine and associated data set into the computer. The computer is used to score the tests and the results are returned to the appropriate team via the teletype network (Figure 4). Turn-around time has been reduced to one or two days, depending on computer availability.

One of the largest problems we encountered in program development was how to best score and analyze the test results. One goal was to determine individual strengths and weaknesses while another goal was to identify and correct deficiencies within the training program itself. We developed a "quick-look" SAT report to meet the first requirement. This report is based on the examinees' results compared to the existing fleet, or total, results contained in the answer file. Individual training recommendations are also contained in the SAT Report and are based on Z-scores greater than 0.5 below the fleet mean. This criteria was picked arbitrarily but has worked well in practice.

A more detailed analysis is conducted to meet the second evaluation goal. This analysis includes individual and group evaluation (i.e., submarine crew, tender crew, instructors, etc.), PPP table evaluation, test instrument evaluation. These evaluations are conducted after each test version is retired and replaced with a new version. The FBM program provides a unique environment for many of these analyses. Community idiosyncrasies include (1) relatively stable crew composition, (2) detailed hardware performance data, (3) close liaison among all activities, and (4) extensive personnel history data. These attributes enable crew performance and empirical test validity evaluations beyond those which are normally practicable.

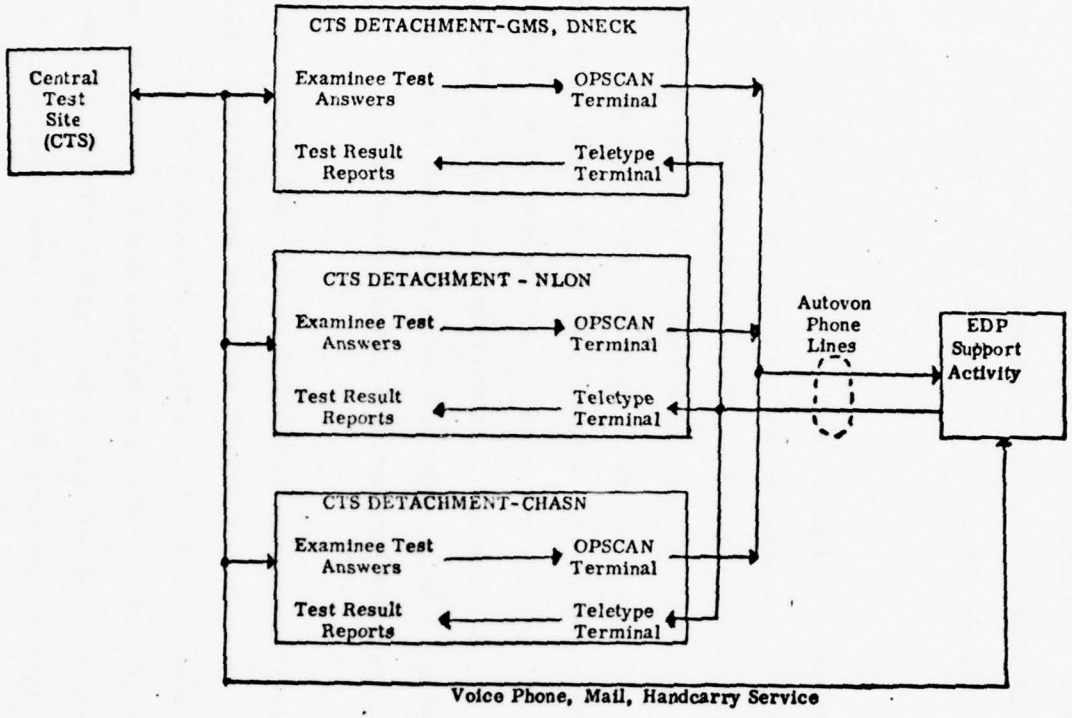


Figure 4. PTEP EDP Communication Network

One fact became readily apparent when crew scores were compared initially. The differences in higher and lower rated technicians in the various crews greatly affected the group overall score. A method was devised to determine an "expected" score for each group based on what the group score would have been, had each examinee scored average for his rate. Actual and expected scores are plotted on crew trend analysis graphs and assist greatly in spotting significant deviations.

During test item analysis, those test items whose characteristics do not meet established discrimination criteria are automatically identified for review and printed out from the computer's test item file. These test items are then reviewed by Central Test Site personnel to determine if they should be deleted or revised, or if they provide an indication of a training problem.

We recognized early in the PTEP developmental phase that extensive computer data processing would be required to handle the quantity of data to be contained within the system. Accordingly, the following five major computer program sub-systems were developed (Figure 5):

1. Test Generation Subsystem. The test generation subsystem includes the programs necessary to produce the tests.
   a. Test reference data file load and maintenance program creates and maintains personnel records.
   b. A personnel report program retrieves and prints out the contents of selected personnel records.
   c. A personnel survey program prints out key information from all personnel records.

BEST AVAILABLE COPY

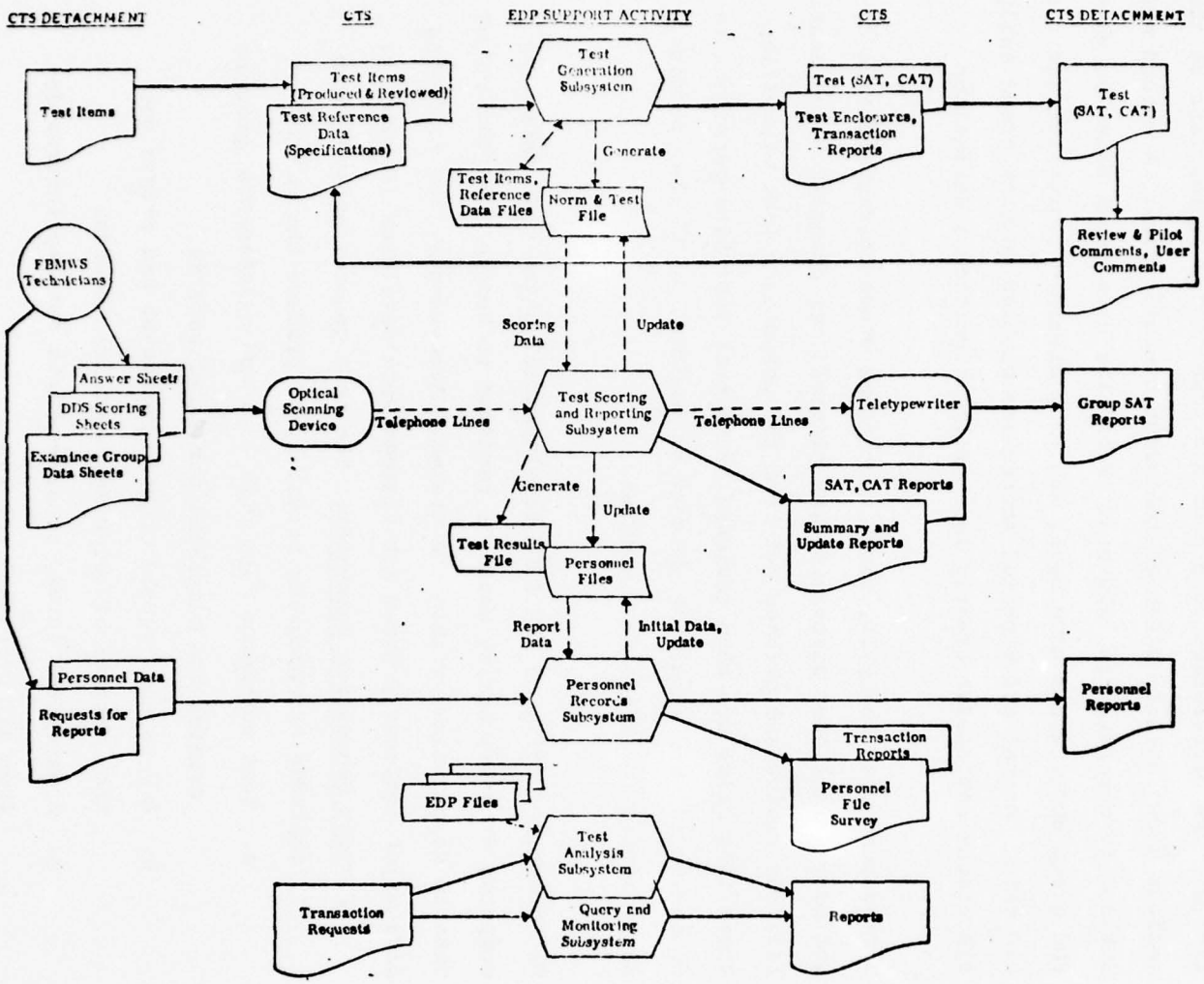


Figure 5 - PTEP ADP SUBSYSTEMS

3. Testing Scoring and Reporting Subsystem. The test scoring and reporting subsystem includes programs for data transmission, for scoring and reporting of test results, and for update of required permanent files.

   a. An input transmission (teleprocessing) program enables remote transmission of test data via optical scanning equipment.

   b. Test scoring and reporting programs assemble test data, score the data using stored answer keys, update fleet norm files, and prepare reports.

   c. An output transmission (teleprocessing) program transmits test result reports back to the originating test sites.

4. Test Analysis Subsystem. The test analysis subsystem manipulates accumulated test results data to provide information used in evaluations of tests, personnel, and training.

   a. Item statistics programs compute attractiveness indices for each test item response, discrimination index for each test item within an overall knowledge test part, and summary difficulty and discrimination indices by knowledge area, TOS level, etc. An engineering review report is also prepared to list those test items failing to meet a specified value of the discrimination index.

   b. A score analysis program computes standard T-scores for each examinee, computes actual and expected average scores for each examinee group, and identifies examinees and groups for later, manual evaluation.

   c. A skill test analysis program recomputes skill exercise scores based on final performance data accumulations and computes summary statistics for use in evaluation activities.

5. Query Subsystem. The query subsystem provides flexible access to the various files to select, retrieve, and report information for evaluation and management.

The majority of problems experienced during development of the ADP support systems were commonplace; the normal growing pains, problems with formats and record contents, and so on. Perhaps the most interesting (and frustrating) set of problems occurred in connection with transmission of test data from the test sites to the computer facility. Our equipment provides remote access via "BTAM", Basic Teleprocessing Access Method. We presumed, as did the vendors with whom we worked, that the remote access would be the least of our problems: you hook the pieces together, write a little program, and everything works. Unfortunately, it did not fall together that easily. If you have been in a similar situation, you probably have experienced the problems, sometimes humerous but more often maddening, that we ran into trying to obtain information, and to purchase and coordinate hookup of equipment from the telephone company, the vendors for the mark-sense reader and the data sets, and the computer manufacturer. It appeared that we were reinventing the wheel. The system finally jelled, and today we have a fairly smooth remote access capability.

Most of the PTEP ADP support programs are fairly typical COBOL applications. One, however, is not; and I will describe it briefly. As data began to accumulate, people naturally began wanting to use it. We had planned to create a subsystem similar to the Management Information Systems so much talked about in the past four or five years.

Unfortunately, we could not precisely define our information requirements: we were accumulating quantities of data to which a seemingly endless number of questions could be addressed.

We settled eventually for a general purpose "query" system. This system can access any computer readable file of fixed-length records, including, of course the PTEP files. It works in three stages, starting with a request for information:

1. A man processes the original question to define a logical procedure for extraction of information from the file, or files, that contain relevant data. His function is essential; he designs the algorithm to extract the desired information.
2. A COBOL program reads the algorithm and writes the implementing program (also in COBOL).
3. The new program, which is unique for the specific question, is compiled and executed, producing a report tailored to answer the original question.

The system capabilities are extended by allowing the output from one question to become the input to a second question. Basically, if we have accumulated the relevant data, and if the question can be resolved to a set of logic equations, the query system can extract the needed information. One major advantage of this system is its flexibility. We can change our file structure around without losing the ability to readily access all data.

Looking to the future, we are working on new analysis techniques which will be required as the CAT program develops. Since these tests are criteria referenced, the traditional difficulty and discrimination formulas contained in our test analysis programs will not apply. Several new types of test instruments are under investigation in the hopes of decreasing the testing time and lowering the cost of acquiring new test vehicles. It is anticipated that the lessons learned in development of the PTEP for the Poseidon program will provide a sound footing for the forthcoming Trident program.

INVENTORY AND ASSESSMENT AT THE UNITED STATES ARMY WAR COLLEGE

By
Quay C. Snyder

As part of the long range curriculum development program at the United States Army War College a new concept has been added. The new concept is improved assessment and feedback to the student of his learning, values, attitudes, abilities and progress. Feedback is essential not only to any student educational endeavor, but it is also essential to the effective functioning of any educational system. Learning cannot occur without some form of feedback.

The long range development program at the War College includes two separate inventory and assessment subprograms: (1) Student inventory and feedback, and (2) The system inventory and feedback. The focus of this paper is on the student inventory and feedback program, a major segment of the inventory and assessment efforts at the War College.

A discussion of student inventory and feedback immediately brings to mind a student "grading system." Traditionally, there has been no formalized grading academic system at the Army War College, the school at the top of the Army school system. The rationale for the absence of a grading system is twofold. First, the USAWC student is a mature professional with 16 to 23 years of commissioned service. He has been selected carefully on the basis of quality of past performances (each year less than 4% of the eligible officers are selected to attend the War College). These students do not require the motivational impetus of a rigorous grading system. The second aspect of the rationale for not having a formalized grading system is that with

the group of high achievers there is obvious and positive correlation between grading with resultant "relative standing" and student anxiety. The War College experience provides an educational opportunity for the student to reflect and learn in an unstressful environment with a minimum amount of student anxiety--hence, no formalized grading system and no relative standing or "honor graduate."

Since the mission of the US Army War College is to educate and since learning cannot be accomplished without feedback, the task is to provide inventory and feedback but at the same time minimize student anxiety. The new inventory and assessment program attempts to provide personal and confidential feedback to the student without the information being part of his official records.

During AY 1973-1974 only 60 of the 228 War College students are members of the experimental group to participate in the new Inventory and Assessment Program. Four faculty members have been assigned to work with the 60 students who will participate in comprehensive personal inventories and professional assessments. The assessment is accomplished through the use of five separate assessment techniques:

a. Baseline Inventory

b. Optional Personal Inventory

c. Professional Self-Assessment

d. Experiential Assessment

e. Value Gained Inventory

The above measuring devices constitute the basic structure of the assessment program. Each is described briefly below.

Baseline Inventory

This is a battery of measures designed to describe for students, developmental coaches, faculty, and the Army the skills, abilities, and characteristics of the senior army officer at a point in time just before he enters the last and highest echelon of formal executive development, the USAWC. The student must be given a fairly comprehensive picture of his own skills, abilities, and characteristics. If he is to be encouraged to develop himself professionally and personally, he needs at least a quasi-rational basis for determining which areas to work on; and in the interests of attenuating anxiety, the student needs to know how he stands in skills, abilities, and characteristics in relation to his student contemporaries.

A second ancillary rationale for the Baseline Inventory relates directly to the stated goal of "tailoring" the educational experience to fit the individual student. The Baseline Inventory provides, in essence, a reconnaissance of the incoming class and of its individual members. The reconnaissance permits--assuming a reasonable degree of pedagogical flexibility--adjustments in the curriculum to meet the measured needs of the class, or of specific individuals. In the case of tailoring for the individual, the Baseline Inventory provides a rational basis for presentation of workshops, and for "stretching" the individual whose knowledge in a given area equals or exceeds that of the faculty.

The methodology used in the Baseline Inventory assesses the individual's (and, when aggregated, the class's) level of competence in a

variety of areas related directly or indirectly to the curriculum. The inventory can answer, for example, a question such as: "Does this student know the field of human relations well enough to write a book, or, are the principles and terminology of the field so unfamiliar that he must exert extra effort to read or discuss the course material?"

The Baseline Inventory can achieve several inventory and assessment objectives. The student, comparing his individual profile to the competency range envelope of the class, has a rational basis for planning his own program for professional and personal development (e.g., attend workshops for "weak" areas). The student's counselor or coach, studying the profiles of his counselees, has a more objective basis for understanding and coaching his counselees as unique individuals. The faculty, studying the group profile and comparing it with baseline measures of previous student classes, has an objective tool for curriculum modifications which recognize difference in classes from year to year. USAWC, studying individual profiles can accomplish (with computer support) several desirable objectives: (1) tailor student committees toward equivalent skill composition; (2) identify students who are weak in a number of areas and require careful coaching programs; (3) identify, for recruitment as permanent faculty or for use as associate instructors, those students who have exceptional strengths in areas directly related to the USAWC curriculum or mission; and finally, (4) provide for the Officer Personnel Directorate a special "assignment potential" profile which indicates those areas where the individual has strong or exceptional competencies (only when expressly desired by the student).

The first Baseline Inventory was mailed to AY 74 students early in the year prior to the time they reported to the College. The inventory was scored and returned to students as "Feedback." Additionally the faculty was provided with pertinent data as a group result of the inventory to obtain a class profile.

There are several recognized shortcomings of the Baseline Inventory instrument. Both are related to the fact that the instrument is a self-inventory--the individual assesses his own levels of skill in areas as defined by the instrument. We can assume that some individuals will intentionally overstate their own capabilities. We can assume also that some individuals will misinterpret or not fully comprehend the implications of the area descriptions. These possibilities are borne in mind when interpreting individual profiles. However, for aggregate profiles, the shortcomings are less significant because errors occur in both directions; e.g., some individuals will understate intentionally their own abilities. The variations also provide the coach with one indicator of the student's self-concept.

Further, while we have thus far (for purposes of illustration) discussed the Baseline Inventory as a single instrument, we should, in reality, consider it as series of related data collection tests and surveys. Initially, we began with the Baseline Inventory designed around curriculum objectives. Over time, we will substitute validated and established measurement instruments for portions of the Baseline Inventory. The Graduate Record Examination (GRE) can be used to assess general academic competency. The Miller Analogies Test can be used for a general measure of problem-solving ability. There are innumerable

tests available for measuring reading comprehension, listening comprehension, and writing skills. And the Educational Testing Service, given clearly-stated learning objectives and supporting course material, can develop a GRE-like comprehensive examination with alternate forms, based on the foundational elements of the USAWC curriculum--ideal for a "before-and-after" measurement.

Optional Personal Inventory (OPI)

The second major assessment task has as its primary objectives the personal development of the individual. Good research on executive development suggests strongly that the point in time near age 40 and two-thirds through a career is a milestone point for contemplation and adjustment/affirmation of personal and professional goals and *modus operandi*. The Optional Personal Inventory (OPI), supported by the coaching program, is designed to assist the student in this difficult introspective task. The methodology of the OPI is far less structured than that of the Baseline Inventory. Whereas the Baseline Inventory is conducted prior to the student's arrival at the USAWC, the Optional Personal Inventory is dependent for its success upon a sound student-coach relationship and cannot commence until such a relationship has been initiated. A battery of inventories are available to students on an optional basis. Included in the battery are instruments to measure attitudes, speaking, reading and writing ability, personality, needs, vocational interests, etc. Initial coaching sessions with incoming students covered the purpose of the inventories and stressed the most appropriate time for making clear to the student that the OPI program is optional and that he can at any time withdraw from the program.

There are two sources of evidence which suggest the strong potential of the OPI concept.

In the executive development field, one of the most promising and rapidly-growing methods of developing executive potential at higher levels is the assessment center. Executives visit these centers and participate in an "assessment week," wherein measures/observations are made of various skills and characteristics which have been found to be closely related to executive proficiency. An intensive "feedback" session follows the assessment week. During the feedback session, the executive is given objective and professional interpretation of the measures made during assessment week. This is privileged information; the staff assessor is not a member of the parent organization; and the individual executive is free to do what he wishes with the feedback information he is given (written and graphic material, audio and videotape cassettes). The US Army investigated the potential of these centers in 1971-72. As a result, more than a dozen general officer designees have attended assessment centers last year; and plans, personnel, and funds for establishment of several Army assessment centers have been approved.

The second support for the OPI concept comes directly from the experience of approximately 15 USAWC students who participated in a Personal Inventory and Analysis program in AY 72. These 15 students were enrolled in a graduate-level counseling program at nearby Shippensburg State College. They all participated in a wide variety of assessment tests and situations; then each took all of the data pertaining to his own performance and did a complete counseling

"work-up" (scoring, interpretation, prognosis, prescription) of the available data. The student group was highly enthusiastic about this in-depth objective exercise in self-development. There was virtually unanimous agreement that it was extremely valuable in terms of personal and professional development.

Professional Self-Assessment

Rationale for this assessment exercise springs directly from the law of learning noted earlier, i.e., without knowledge of results, attempts to learn are only random behavior. Therefore the purpose of professional self-assessment is to individualize feedback to the student regarding his attempts to learn, and his personal, professional and leadership development; to provide feedback to the instructor regarding his teaching; and to provide a rich source of empirical data for continued curriculum improvement.

Essentially, in the professional self-assessment, the methodology requires no change in the present group requirement which accompanies each course. In addition to the group requirement, however, there should be an individual requirement. At or near the end of the course, the student is given an examination which contains subjective questions requiring responses at levels of knowledge and in content areas derived from the stated course objectives.

The student is given perhaps two hours to address all of the questions. His answers can be in any format: written, outlined, topical, or graphical. At the end of this time, he is provided with a solution. This is not a "school solution," but rather, an "expert solution"--prepared by a faculty member who is an expert in the field,

and who, if presented with the examination questions in an official capacity, would respond as indicated on the solution sheet.

The student, by self-assessment, compares his answers with those on the solution sheet. He is then given the opportunity to discuss, debate, or support the components of the expert solution--preferably with the expert himself.

From this comparison and discussion process there undoubtedly will come valid individual minority views and creative comment. These should be encouraged (but not in class) and can serve as a basis for subsequent individualized discussion between student and faculty.

At the completion of the examination-comparison-discussion process, the student turns his paper in to the faculty member. There is a clear stipulation that no paper be identified with the individual student's name. (An "option" on signing examination papers would have coercive implications). The purpose of this turn-in of examination papers is to provide the faculty with empirical subjective feedback on how well learning objectives were achieved.

Experiential Assessment

Experiential assessment consists of three separate assessment tasks conducted while the student is devoting his entire educational effort to participation in a group problem-solving exercise. The assessment for this phase is termed "experiential assessment" because the research objective and the group problem-solving method represent a model of the professional requirements that the student will most

probably face following USAWC graduation. With some exceptions, the professional requirements which he is being prepared to meet will involve working, for a superordinate leader or director, as a member of a group (staff or committee), and producing a product which must withstand the scrutiny of experts in the field. The three assessment tasks of this phase are geared to the assessments the student will encounter in his post-graduate environment; i.e., those of his superior, his peers, and his critics. The team leader or project director, prepares an assessment of the individual student which would be somewhat similar to the Officer Efficiency Report.

A second task of experiential assessment seeks to provide the student with feedback pertinent to his performance as a member of a team as seen by his colleagues or peers. Again, the policy of assured anonymity will apply, and the student himself is to be the sole customer for the peer appraisal information.

The third task of experiential assessment is appraisal by subject area experts. This third assessment task has its major impact in the motivational area rather than the developmental area. From the outset of the project, research team members will know that their group project is to be evaluated by an outside expert in the field, engaged specifically to function as a critic of the group research product.

Feedback from the three assessments of the Experiential Assessment task should provide the student with clear and detailed insight of what he should do to function most effectively in the problem-solving environment he is most likely to face after graduation.

Value Gained Inventory

We have good evidence that change occurs within the student during the USAWC year. We assume that changes occur also in other areas--in functional, academic, geographic, and interpersonal abilities, for example. But we are hard-pressed to say just what has changed, and how much, and in what direction.

The assessment task of the Value Gained Inventory is an attempt to determine the total developmental effect (upon the student) of the year spent at the US Army War College. The student needs this information in order to review and appreciate the professional and personal development brought about by the total educational experience. Such a review appears prerequisite to individual planning for continuing development, extending far beyond USAWC graduation.

The Value Gained Inventory is even more critical to the War College itself. We need a year-end summary assessment to determine the degree to which our educational system has achieved its curricular objectives. We need facts to answer with some precision the often-asked question, "What sort of man does the USAWC produce?"

The methodology of the Value Gained Inventory consists simply of administering alternate and equivalent forms of the instruments used at the beginning of the academic year, would be sufficient time for the data collection requirements of the Value Gained Inventory. Most of the instruments will be components of the Baseline Inventory, but measures from the Optional Personal Inventory could be included.

Statistically significant differences, on pairings of data for the same individuals, strongly suggest effects directly attributable to the USAWC year.

Summary

The program of student inventory, assessment is a trial program conducted on an experiemtnal basis for Academic Year 1973-1974. If after a comprehensive evaluation, the program is determined to be worthwhile, it may be modified to include a larger segment of the students at the Army War College. Those portions of the program determined to be of marginal value will be eliminated.

During the conduct of the program certain basic principles must be followed. First, the student participants must realize the program is oriented toward helping them sharpen the edges of their own skills, learning, values, and knowledge. They are not considered patients or unhealthy individuals, but successful, mature, proven executives. An analogy can be made of their career performance with that of a marksman rifle range. Thus far in their career, they have already fired "expert." But until this time they have not always known where their shot groups have been located on the targets. The faculty coach has the function of marking some of their targets. Perhaps with a minimum amount of coaching the shot group can be moved to the center of the bullseye. With the proper coaching even an expert can improve his "score."

The second principle is that of confidentiality. The student participant must be assured that none of the data gathered will be attributed to him personally. The information must not be used in any

official records or reports which identify the individual. The relationship between the student participant and the faculty coach must be a privileged relationship.

The other principles e.g., honesty, empathy, respect, concreteness, self-disclosure, and immediacy, of an effective helping program are also important. Since the program is on a trial basis this year there is some skepticism and apprehension by a few of the observers. For the most part, everyone participating in the inventory, assessment and feedback effort is enthused and confident it will enhance the educational endeavors at the United States Army War College.

SUBORDINATE RATINGS: WHY NOT?

W. H. Githens and R. S. Elster

U.S. Naval Postgraduate School

The notion of having subordinates rate their seniors is one that typically engenders a great amount of feeling. These authors became interested in the topic while teaching graduate level psychology courses to military officers when we found that a quick way to provoke class discussion was to assert that a system for gathering subordinate evaluations of seniors seemed deserving of experimentation. Over the last several years, therefore, we and a number of our students have conducted some studies concerning subordinate evaluations. It is to these studies that we will now turn.

TWO GENERAL SURVEYS CONCERNING SUBORDINATE RATINGS

In 1971, Lt. J. G. Bloomer, USN, working with one of the authors, sent 1100 survey questionnaires to the U. S. officers attending the Naval Postgraduate School. Over 350 officers responded to this survey concerning subordinate evaluation. Lt. Bloomers' survey focused on the acceptance of subordinate evaluations when they are for the senior's use only. Thus, his instructions read, in part, "...The purpose of this survey is to discover whether the students of NPS feel that the military officer might be benefited by a program such as this, in which he is evaluated by his subordinates, with the ratings submitted directly to him solely for his own use. The responses obtained to this survey are summarized in Table I.

TABLE I

Results of Lt. John Bloomer's Survey
of Naval Postgraduate Students
Attitudes Concerning Subordinate
Evaluations (N=350)

1. Do you feel such a periodic survey by his enlisted men would be a significant help to the junior officer?

Yes _____ Maybe _____ If no, go to part III.

Note: Lt. Bloomer grouped the number answering "Yes and "Maybe" into "those finding some merit in the program," since in the majority

of cases the "Maybe" answer seemed to indicate concern with the mechanics of the program, not necessarily the value.

Total finding some merit in the program:

Overall (all officers responding)	74.5%
LT and below	83.2%
LCDR and above	68.0%
Army (of 10 responses)	80.0%
Marines (of 22 responses)	45.5%

2. Should such a program be restricted to the junior officer only?

Yes 15.4% No 83.6%

Note: The most common response to "level it should be restricted to" was all below flag rank; however, many felt such feedback would be helpful all the way to CNO.

3. Do you feel that such a system would have a significant effect on enlisted morale?

Yes 61.5% No 15.8% Don't know 22.7%

4. If you were rated by your subordinates, would you consider these reports:

Quite seriously 33.4% Casually 3.9%

90.8%

Seriously 57.4% Other 5.3%

5. If you found such reports adverse, would you attempt any changes in your leadership techniques?

Yes 72.7% No 0 Don't know 27.3%

Note: This question elicited many comments, and in most cases the "Don't know" answer was a qualified "Yes". That is, after consideration of the source, how the program was run, whether they were isolated reports or a trend, etc., a leadership style change might be attempted.

6. Do you feel that Commanding Officers might gain by having a similar evaluation made of them by their junior officers?

Yes 76.4% No 14.0% Don't know 9.6%

Note: It is interesting to note that many who had held command commented that this feedback would have been especially useful to them.

7. Do you feel that such feedback might have been of help to *you at any time in your career?* If yes, when?

Yes 87.9% No 12.1%

Note: The "when" in this case was generally as a LT and below, but many responded that it would help at any time.

8. Do you feel that enlisted raters would conscientiously attempt an accurate evaluation of their superior officer?

Yes 74.2% No 7.2% Don't know 17.6%

9. If such an evaluation system were adopted by the military services, do you feel that the raters should be restricted to:

No restrictions (all levels of subordinates) 58.3%

Petty Officers only 34.6%

Other 7.1%

Note: The "Other" in this case generally was accompanied by comments suggesting a graded grading system; that is, officers and CPO's grade CO & XO while Chiefs and First Class rate their Department Head, or some similar groupings. There is certainly some merit to this idea, if only to reduce the amount of paperwork.

The results of this survey appeared to indicate two points conclusively: We quote Lt. Bloomer, "(1) There is a general concern for improving officer/enlisted lines of communication. (2) The overwhelming majority of respondents indicated a high degree of respect for the judgement of the enlisted man of today." The reader should keep firmly in mind, however, that Lt. Bloomer's survey had as a major premise that the subordinates evaluations would be for the use of only their senior.

At about the same time, Lt. Bloomer was conducting his survey, Githens surveyed 68 USN officers at the Navy Postgraduate School. Two of his questions involved the issue of ratings by subordinates. The questions and the responses obtained are presented in Table II. The reader should notice that in the instructions to Githen's survey, unlike those to Lt. Bloomer's survey, the use of the data from subordinates ratings were not explicitly restricted to use only by the senior.

TABLE II

Results of a Survey by W. H. Githens of 68 USN Officer/Students at the U.S. Naval Postgraduate School

As part of an overall evaluation, what do you think about being rated by your subordinates?

An excellent idea	3
A good idea	13
A fair idea	17
A poor idea	34

Assuming that peer, subordinate, self, and the present "rating by superior" were in operation, what should be their contributions to the total evaluation? Assign a weight from 0 to 100 each system. The sum of the weights must add up to 100 units.

Rating by peers	13.87*
Rating by subordinates	8.00*
Rating by superiors	72.23*
Rating by self	5.79*

* Means computed over sample of 68 officer/students.

The results of Githen's survey seem to show that officers are less sanguine about subordinate ratings than did the results of Lt. Bloomer's study. We feel that this difference in results stems from the psychological sets given by the questions and instructions to the surveys. Lt. Bloomer's survey addressed using subordinate ratings as a feedback vehicle for the cognizant senior, while Githen's survey asked the respondents about using data from subordinate ratings when evaluating officers. These results lead to the rather obvious conclusion that the use to which subordinate evaluations are put will influence the reception they receive from those who are evaluated. Thus, it is necessary to understand the fears and concerns that officers have concerning subordinate ratings.

In his survey, Lt. Bloomer asked those who were opposed to the type of subordinate rating system he proposed to comment on their reservations. Table III contains a representative set of the comments that were obtained.

TABLE III

Representative Comments of Officers
Opposed to a Subordinate Rating System
as proposed by Lt. Bloomer

LT	"This system of evaluation would turn into a popularity contest between junior officers." Note: This comment was the most common reason given for the negative response.
LT	"I believe face-to-face discussions are more productive."
CDR	"Undermines discipline. A crutch for poor leaders who don't have the ability to communicate with men, or to accurately comprehend what is going on around them."
LT	"It is difficult enough aboard ship to instill in a junior officer that the principles of good management are best served by a professional relationship with his men rather than a friendship basis. This evaluation would only serve to promote the latter."
LCDR	"There is just too much danger in the proposal for all concerned for the few benefits that might be realized."
LCDR	"A good officer knows how his men feel about him. The officer who does not know is unlikely to be affected, pro or con, by written evaluations."
LCDR	"There is enough breakdown in authority in the modern Navy without open criticism by subordinates."
LCDR	"I feel peer group pressure or influence would greatly prejudice the non-rated opinion. Senior Petty Officers would be more capable of fair evaluations."
LT	"The main reason I said no, is that I feel the evaluation would be degraded to a grievance session..."
LCDR	"Personally, such a system in the military service goes against my grain."
LCDR	"God, when is this liberalism ever going to cease?"
LT	"Once comments are noted on paper, the 'Navy way' ensures the CO/XO, Dept. will manage to see them."
LT	"Strongly feel that feedback should be confidential and not available to superiors."

LCDR "I feel a junior officer is very formable in his early years and should be guided only by his seniors, not his juniors."

LCDR "The evaluation would seem to put too much emphasis on molding the junior officer to suit the needs of subordinates, when in fact the molding should come from the top -- providing of course that the superior is capable himself and takes advantage of opportunities to observe the junior officer."

LCDR "One would tend to cater to the wishes and whims of subordinates if only to protect his ego when he receives their 'biased' evaluations."

LCDR "Feedback should be welcome in any form. It enables one to see if points of emphasis are coming across. It should be entirely optional on the part of the enlisted men and should not go through any chain of command, i.e., leading PO's or Chiefs."

LCDR "The typical junior officer is too sensitive to the opinions of his subordinates. By trying to please all of them, having no way to differentiate between serious comment and sarcastic, he could ruin himself in his own image. He could lose whatever self-confidence he had built up."

LCDR "I feel the success of such feedback would depend upon whether the recipient could take criticism, and if he could, effect changes where necessary, There is a danger of an impersonal system such as that described becoming a routine, meaningless exercise if the recipient cannot communicate its worth by viewing the feedback seriously and making changes where necessary."

LT "Problem in maintaining anonymity comes up when the officer has very few subordinates. May need to propose a system in which evaluations are only made when the officer has, for example, five or more subordinates." Note: Some form of this comment was often made and it is probably true that maintaining true anonymity in a small unit would be difficult, if not impossible.

CDR "The evaluation should not be forced upon the subordinates. It should be optional whether or not you evaluate your senior."

LT "Since seniors have a hard time evaluating juniors and our fitness reports are a highly controversial topic, don't you think an uneducated individual would have even a more difficult time efficiently evaluating seniors?'

LCDR "I don't feel a PO is qualified to tell me how to do my job!"

LT "I do not believe that enlisted men, in general, have the background to evaluate my performance, since personnel management is only part of my job as a Naval Officer. Most could not evaluate my shiphandling ability since most have no experience. They could not evaluate my material and financial management since they do not have the training I have, etc., etc.

LT "I do not feel most younger PO's and non-rated men would be honest and objective in making out such evaluations. Also, I feel it would be difficult to keep these evaluations from the concerned officer's superiors."

LT "I believe that the average sailor working for a young JO is not mature enough to make the type of comments which would be beneficial."

LCDR "I think the whole idea is a bunch of ----! If you can subject the EM's to the same pressures, hold them to the same responsibilities, etc., THEN they are in a position to rate him objectively, and not before."

LCDR "A good manager should be able to evaluate his policies by the results of his efforts or by another means of feedback. Subordinate evaluations are a nice thought but probably useful only if the superior has really lost contact with his subordinates. If this has happeded, chances are the evaluation won't help matters."

LT "The junior officer who is concerned enough to care what his subordinates think of him does not need such a device."

LCDR "Any JO who does not get sufficient feedback in this area is not worth his salt anyway. Sounds like just another poor attempt at leadership by the numbers."

Perhaps the major theme that can be gleaned from examining the comments in Table III, is that officers opposed to subordinate ratings perceive favorable subordinate ratings as resulting from the officer conducting himself as one super-nice guy who is lenient and lax with his subordinates. In the eyes of many of the officers opposed to such a rating program, a subordinate rating system would promote a popularity contest.

Without arguing the relative merits and demerits of leader popularity per se, one can turn to the issue of how do subordinates, in fact, develop their opinions of their seniors. That is, what behavior of officers are are deemed effective and ineffective by their subordinates? The authors and a number of their students have conducted two studies in attempts to understand what leads an enlisted man to label an officer as "effective" or as "ineffective".

Two Critical Incident Studies Concerning Subordinates' Views of Officer Effectiveness

Both of the studies to be discussed used the critical incident methodology (Flanagan, 1954) to gather responses from enlisted men concerning effective and ineffective officer performance.

The first study used enlisted personnel assigned to two Navy Jet Aircraft Attack Squadrons. One hundred six men representing pay grades E-2 through E-9 were included in the sample. The questionnaire form asked the respondent to describe an officer's behavior in the situation during your Navy experience when you considered an action on the part of the officer for whom you worked to be the best example of an effective Naval officer." Essentially, the same question was then later presented, asking the respondent to describe an ineffective Naval officer. The reader should notice that the questions address effectiveness and ineffectiveness in rather broad terms, and foci such as unit goals or mission accomplishment were not provided. These questions to the respondents should not have served to preclude responses concerning, say, discipline by the officer, or the lack of discipline by the officer.

Table IV presents a summary of the critical incident categories formed from the responses to the two questions. (The original content analysis yielded a number of more specific categories. These specific categories were then combined into the behaviorally richer categories shown in Table IV.)

TABLE IV

Broad Categories of Responses Made
Concerning Effective and Ineffective
Naval Officers - Study I
(N=106 Navy Enlisted Men)

Effective Naval Officer		Ineffective Naval Officer
54	Sensitivity to Human Needs	34
20	Professional Competence	29

Effective Naval Officer		Ineffective Naval Officer
9	Trust in Subordinates	13
5	Involvement in Job	
	Personal Characteristics	7
18	No Response	23
106 respondents		106 respondents

To give the reader a flavor of the incidents obtained and categorized, the following paragraphs list some excerpts of representative incidents.

Sensitivity to Human Needs, Effective:

> An enlisted man was having problems with drugs; this CO set up a drug abuse program which really helped this man and some others.

Sensitivity to Human Needs, Ineffective:

> After a plane crash in which the pilot was killed, this officer commented, "men can be replaced".

Professional Competence, Effective:

> Ship was anchored in Hong Kong when action had to be taken due to a typhoon. The LCDR took charge and secured the ship while half the personnel were off on liberty.

Professional Competence, Ineffective:

> One of the pilots didn't put his hands up so we could de-arm his guns. He by-passed the de-arm area and pulled his plane into the line to stop with his guns pointing at the ready service magazine.

Trust in Subordinate, Effective:

> The department head gave me the assignment and left its completion up to me.

Trust in Subordinate, Ineffective:

> My division officer caused me to lose a launch by insisting that the gyros on an air-to-air missile wouldn't be uncaged when power was applied to the plane, even though all of us in the shop were experienced and told him he was wrong. He let us know he didn't feel he could believe us.

Involvement in Job, Effective:

> During carrier qualifications, pilots in three planes in a row held brakes while going off the catapaults. The division officer saw we needed help fixing the tires so he pitched in to get the job done.

Personal Characteristics, Ineffective:

> In my last squadron, my CO was a heavy drinker and would carry-on about things. This negatively influenced the morale of the men.

The authors of this paper and the three Navy officer/students who worked on this study noticed one other thing about the incidents that had been gathered; none of them seem petty and whiny. Most of the incidents seemed to bore in on behaviors that the Navy would want to reward or extinguish in its officer corps. Further, the sort of content areas derived from the incidents look like facets of performance the Navy may wish to consider when evaluating officer performance.

While the research just described was being conducted, several other officer/students collaborated with the present authors in another critical incident study. This second study used a questionnaire which was answered by 289 Navy enlisted men in pay grades E-6 through E-9, who were at five Fleet Training and three Air Technical Training Centers. The questions asking for critical incidents were somewhat different from the ones in the first study, as they explicitly mentioned mission accomplishment. The question concerning an effective critical incident asked: "Think of a specific time (date and place) when an officer, whom you worked directly for, did something that contributed directly to the successful accomplishment of your units mission. Exactly what did this person do that was helpful

to you or other persons in the branch/division/department?" The critical incident question addressing ineffective performance was similar to the one above, but asked about something that an officer did that actually delayed or hindered accomplishment of the units' mission.

The content analysis of the 269 critical incidents received in this second study was conducted by a group of officer/students working independently from the group who content analyzed the data from the first critical incident study. The results of the second critical incident study are summarized in Table V. (Where it would not do violence to the categories found in the second study, their labels were changed to match those found in the first study.)

TABLE V

Broad Categories of Responses Made Concerning
Effective and Ineffective Naval Officer
Performance - Study II
(N=289 Navy Enlisted Men)

Category	Percent of the Total Responses in this Category
1. Trust in Subordinates	26.8
2. Professional Competence	24.3
3. Involvement in Job	13.4
4. Sensitivity to Human Needs	11.4
5. Training of Subordinates	10.4
6. Communications	6.5
7. Safety	5.0
8. Discipline	1.5
9. Over Familiarity with Subordinates	0.7

A comparison of Tables IV and V shows that there is a high overlap between the categories derived from the critical incidents gathered during the two studies. The second study (Table V) yielded five categories not found in the first study (Table IV): Training of Subordinates, Communications, Safety, Discipline, and Familiarity with Subordinates. The first study, on the other hand, yielded a Personal Characteristics category not found in the second study. The authors suspect that the difference between the questions in the two surveys -- the second addressing mission success while the first

did not -- was the major reason the categories found from the two studies did not map even better onto one another.

Two categories in Table V probably require some additional comments, as they might seem to substantiate some of the reservations held concerning subordinate ratings; these are the categories labeled, Discipline, and Over Familiarity with Subordinates. The incidents under Discipline typically mentioned punishment of a group. For instance, "Punishment was given to the whole radio gang by rescinding special liberty when one individual was at fault for not delivering a message," was one of the incidents in this category. Incidents in the category labeled Over Familiarity with Subordinates, all referred to cases in which junior officers bypassed senior enlisted men in the chain of command when dealing with enlisted men. The individual reporting such an incident apparently felt the chain of command was bypassed because the junior officer was too familiar with some of the junior enlisted men, and that this hindered mission accomplishment.

As was the case of the first study, none of the critical incidents gathered during the second study seemed of a petty nature. Instead, they usually addressed performances that the Navy would either wish to reward or extinguish. Neither study provided support for the notion that subordinate ratings would be greatly influenced by officers adopting strategies of ingratiation in dealing with their subordinates.

Discussion

The use of subordinate ratings has been discussed for many, many years. Based on armchair analyses of the topic, it has been pointed out that subordinates are in a position to observe some of the performance of their supervisors. In many cases the subordinate is in a unique position which affords him opportunity to observe supervisory performance that is inaccessible to others. Assuming that more information available concerning various aspects of a supervisors performance will permit a better evaluation of his over all performance, it follows that subordinate ratings should be used. But even though this was all pointed out long ago, subordinate ratings have not been utilized in practice. This non-use is evidently based on the conceptions of a "popularity game", limited perspective, etc.

The most significant finding of our critical incident studies in this area, is that the bases used by subordinates for their evaluations of their superiors were not consistent with such conceptions. They were, for the most part, based on aspects on which most supervisors would want to be rated. The naval officers, who as part of their academic training, gathered the critical incident information for us were originally suspicious of, fearful of, and generally opposed to, subordinate ratings prior to their work in this area. Their switch in attitude after having gathered and analysed the critical incident information was dramatic.

Subordinate ratings, of a sort, have recently received an upsurge in use in academia. Students' evaluations of teachers or professors are no longer an uncommon practice and at least one firm, the Apex Corporation, uses subordinate ratings. The main resistance to their use seems to be based on possible misconceptions about what the subordinates would actually be rating. If this is so, what information could be generated to correct these possible misconceptions? Would it be necessary to have all supervisors do studies in which they gathered critical incidents? With our recent renewed attention to the "people" aspect of organizations, subordinate ratings may be more important and possible now than ever before. Since our critical incident studies argue strongly against the reasons usually given for not having subordinate ratings, the question should again be asked: Subordinate Ratings: Why Not?

References

Flanagan, J. The Critical Incident Technique, Psychological Bulletin, 1954, 51, 327-358.

Lutrell, H. Performance Appraisal at Apex Corporation: Discovery -- Or Delusion? National Business, March 1972.

The Keesler Study - Electronic Technicians Four Year Evaluation of Three Types of Training

Virginia Zachert, Ph.D. Consultant
Medical College of Georgia
Augusta, Georgia

(For the 1973 Military Testing Association Conference
San Antonio, Texas, 31 October 1973)

INTRODUCTION

A Department of Defense service test program was established in 1965 at Keesler Air Force Base, Mississippi, to explore ways to reduce the time and cost of training non-prior service airmen to perform as electronic equipment repairmen. Two occupational specialists and two markedly different approaches to training were compared with a control group. During the year 1966, 441 graduates of the four experimental courses were matched with 294 control graduates of corresponding regular courses and assigned to field units in four commands. Job performance data were collected throughout their first active duty tour and analyzed to compare their success on the job.

MAJOR EVALUATION STEPS

There were five major evaluation steps as follows:

After-Course Proficiency Test at Keesler Techinical Training Center the first week following graduation. (Appendix A)

First-Job Proficiency Evaluation - first ten weeks on the job. (Appendix B)

Delayed Field Evaluation - nine months after completion of resident training. (Appendix C)

Second-Job Proficiency Evaluation - after approximately two years on the job. (Appendix D)

First Enlistment Termination Survey - at completion of initial active duty tour. (Appendix E)

TECHNIQUES USED

The following techniques were used to gather evaluation data:

Performance and written tests administered to graduates.
Job Proficiency appraisals by field supervisors.
Mailed questionnaires completed by graduates and their supervisors.
Field visits by evaluation team members.
Survey of records at Air Reserve Personnel Center.

FINDINGS

I. After over one year of graduate field experience, findings indicated that:

According to graduates and their supervisors, training provided in the control (regular) courses is very closely aligned with the needs of field units and is compatible with established unit OJT programs, career development courses, and specialty knowledge testing requirements. Field supervisors almost unanimously oppose any reduction in the current scope and quality of resident technical training.

The shorter service test courses with drastically reduced electronics fundamentals training ("X" course), while not completely unsuccessful, resulted in several important training deficiencies which contributed to initial and lingering graduate job difficulties. Most apparent among these were:

Too limited a technical vocabulary resulting in faulty communications.

Inability to apply, to job needs, an understanding of elementary electronic phenomena such as capacity, bias, resonance, transients, inductance, ets.

Inability to interpret detailed circuit diagrams.

Insufficient knowledge of the basic functioning of classic electronic circuits.

Too limited an ability to troubleshoot (theoretically and practically) electronic equipment at the circuit component level. Field supervisors had to give compulsory, compensatory electronic fundamentals training to graduates of these abbreviated training programs. Graduates were very vocal in their adverse criticism of the "X" type experimental training they received.

The shorter service test courses, which emphasized electronic fundamentals training but de-emphasized specific equipment training, ("Y" courses) were more successful than the "X" courses. However, the graduates did encounter substantial difficulties among these deficiencies in graduate hob proficiency were:

Lack of sufficient initial skill in using published technical data (technical orders, maintenance manuals, etc.)

Too little versatility and skill in using standard test equipment.

Too limited a knowledge of electronic systems operation and maintenance techniques (this was very pronounced for aircraft equipment repairmen because they were responsible for maintaining several different types of equipment even during initial job assignments). Graduates were also very critical of this training.

Evidence to date, although incomplete, indicates that if field conditions remain the same adoption of the shorter courses would:

Reduce initial job proficiency.

Reflect additional training workload and loss of productive work into operational units.

Under present standards, increase SKT failures.

Require more time on the job to upgrade to the "5" skill level.

Increase the percentage of graduates believing their training to have been inadequate from 3-11 percent to 33-60 percent.

II. After approximately three years of the Service Test Program, findings indicated that:

The experimental graduates compared favorably with the control graduates after three years of job experience and considerable formal training and job oriented training.

Experimental graduates and supervisors were still critical of the abbreviated type resident technical training received by the experimental graduates.

If field conditions remain the same, adoption of the shorter courses would apparently be the same as stated earlier.

INFERENCES THAT CAN BE MADE

The service test was conducted essentially as planned; however, there were some difficulties occasioned by the following unexpected developments: The intensification of the Southeastern Asia Conflict, several successive early-out programs affecting the experimental airmen, and the requirement for an earlier-than-expected completion of the service test.

Considerably more data were collected than could be reduced and analyzed with the time and resources available. However, all data have been carefully recorded in ten main report volumes. Additional raw data have been preserved in the forms of completed questionnaires, data reduction worksheets, recorded comments by graduates and their supervisors, etc.

Training received by the control students was very closely aligned to the needs of field units in terms of graduate job proficiency and career development.

The job proficiency and career development of graduates of both concepts of experimental training fell short of the needs of field units; however, the applied electronic ("Y") graduates were the more satisfactory.

Both categories of experimental graduates required substantially more on-the-job training and supervisory attention in the operational units than did the control graduates. Recorded comments of field supervisors and graduates were consistently critical of the training deficiencies in the experimental courses. The criticism was especially strong for the "X" concept of training.

After about three years into their first active duty tour, 84% of the service test graduate indicated their intent not to extend their active duty commitment. This was generally borne out by a terminal survey of available records in the Air Reserve Personnel Center (ARPC) which showed that 66% of the graduates had in fact terminated their first active duty tour in four years or less.

Research Problem Areas:

1. Select sample for diads and triads.
2. Assign of control and experiment graduates.
3. Maintaining diads and triads.
4. Getting questionnaire data.
5. Analyzing data.
6. Presenting findings.

Appendix A
Sample of After Course Proficiency Test (Performance)

NAME _____ AFSC _____

CLASS _____ DATE _____ TEAM _____

DO NOT WRITE IN SPACE BELOW

OPERATIONAL - PERFORMANCE (0-25-20)
AN/APX-25 -- VOLTAGE CHECKS

TIME REQUIREMENT: You are allowed 20 minutes to complete this exercise. Call the evaluator when you finish or find that you cannot proceed without assistance.

EXERCISE: You are to perform voltage checks at designated circuit points and compare them with the voltage values recorded in the technical manual. The voltage checks will be performed on the RT-279/APX Power Supply and Video Decoder stages.

/ / / / / CAUTION / / / / /

SEE FIGURE 7-54 IN THE T.O. TO LOCATE HIGH VOLTAGE AREAS

/ / / / / CAUTION / / / / /

PROCEDURE:

1. Perform the voltage checks in accordance with procedures outlined in Figure 7-55, Section VII of the technical manual and supply information for the following chart as you proceed.

VOLTAGE CHECKPOINT	VOLTS INDICATED IN T.O.	MEASURED VOLTS
a. V-117, pin 1		
b. V-117, pin 3		
c. V-117, pin 5		
d. V-117, pin 7		
e. V-117, pin 2		
f. V-104, pin 5		
g. V-105, pin 2		
h. V-106, pin 1		
i. V-106, pin 6		

2. Record the serial number of the AN/PSM-6 used. _____

CALL THE EVALUATOR WHEN YOU FINISH

REFERENCES:

1. T.O. 12P4-2APX25-2, Section VII, Figure 7-54
2. T.O. 12P4-2APX25-2, Section VII, Figure 7-55

S_____
F_____
T_____
COMPLETION
YES___ NO___

BEST AVAILABLE COPY

Appendix A
Sample of After Course Proficiency Test (Written)

NAME _____ AFSC _____

CLASS _____ DATE _____ TEAM _____

DO NOT WRITE IN SPACE BELOW

WRITTEN - PERFORMANCE (W-25-1)
AN/APX-25

1. A bench check of a KY-95A/APX-25 revealed that each MODE 3, EMERGENCY reply consisted of two and one-half code trains. Voltage and resistance checks revealed Relay K-_____ was open.

S _____
F _____
T _____

2. A bench check of an RT-279/APX revealed that the AOC/AVC Test Point (X-205, Pin 3 or Test Point T) measured +100 VDC. Further troubleshooting revealed Electron Tube V-_____ was defective.

3. A bench check of an RT-279A/APX revealed -250 VDC at Test Point Y-1 and -48 VDC at Test Point Y-4. Resistance checks revealed Potentiometer R-_____ was open.

4. A bench check of an RT-279A/APX revealed that the transmitter output pulses were 1.7 microseconds in duration. Waveform checks revealed that Delay Line Z-_____ was defective.

5. A bench check of a KY-95A/APX-25 revealed that the Delay Line Z-1601A was oscillating, thus producing random code pulses out. Waveform checks revealed a full sinewave input to the Delay Line at Test Point H (ORN 1). Resistance checks revealed that Crystal Diode CR-_____ was open.

6. When Crystal CR-401, in the RT-279A/APX, burns out and is replaced, Electron Tube V-_____ should also be replaced as a preventive maintenance step.

7. A bench check of an RT-279A/APX revealed that the MODE 2 and I/P replies both consisted of three (3) pulses, 16 microseconds apart. Capacitor C-_____ should be adjusted to correct this malfunction.

8. A bench check of an RT-279A/APX revealed that MODE 3 was inoperative. All other MODES are functioning properly. A voltage check at Pin 5 of Electron Tube V-106 revealed 0 (ZERO) volts. The probable cause of this malfunction is Switch S-_____ open.

9. Blower motor noise is interrogating an RT-279A/APX, causing spurious replies. The first stage to be checked in order to isolate this malfunction is V-_____.

10. A bench check of a KY-95A/APX-25 revealed output pulses 0.2 microseconds in duration. Troubleshooting revealed Delay Line Z-_____ was defective.

1 _____
2 _____
3 _____
4 _____
5 _____
6 _____
7 _____
8 _____
9 _____
10 _____

DAILY PERFORMANCE RATING

RATING KEY

FREQUENCY - Number of times performed

PERFORMANCE - Performance rating (0 - Unfamiliar w/job; 1 - Slight familiarity; 2 - Generally familiar; 3 - Competent; 4 - Skilled)

TIME - Time rating (Senior level comparison) (0 - Unable to perform job; V - Very slow; S - Slow; A - Average; F - Faster than average)

JOB ELEMENTS	TYPE OF EQUIPMENT (When applicable)	1			2			3			4			5			6			REMARKS (IDENTIFY SPECIFIC DEFICIENCIES)
		FREQUENCY	PERFORMANCE	TIME	FREQUENCY	PERFORMANCE	TIME	FREQUENCY	PERFORMANCE	TIME	FREQUENCY	PERFORMANCE	TIME	FREQUENCY	PERFORMANCE	TIME	FREQUENCY	PERFORMANCE	TIME	
67. Receiver Stages		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
68. R. F. stage																				
69. I. F. stage																				
70. Audio stage																				
71. BFO/VFO																				
72. Video stage																				
73. AVC stage																				
74. AGC stage																				
75. AFC																				
76. Coding/decoding network																				
77. Coincidence stage																				
78. Delay line																				
79. Multivibrator																				
80. Oscillator stage																				
81. Mixer																				
82. Detector																				

KEESLER TC OFFICE FORM MAY 65 140a

BEST AVAILABLE COPY

N-X=137 N-C=137

3. ON-THE-JOB TRAINING:

a. Do you feel that technical training at Keesler adequately prepared you for participation in your unit's CDC program? ☐ yes, ☐ no. If your answer is no, explain.*

Yes-X=55; C=123 No-X=82; C=14

b. Did the technical training you received at Keesler provide you with an adequate foundation to further prepare on-the-job for the SKT? ☐ yes, ☐ no. If your answer is no, explain.*

Yes-X=47; C=125 No-X=89; C=12 Omissions - X=1 C=0

Yes-X=83; C=80 No-X=53; C=56 Omissions - X=1 C=1

c. Have you taken the "5" level SKT? ☐ yes, ☐ no. If yes, did you encounter any particular difficulty? ☐ yes, ☐ no. If yes, explain.*

Yes-X=67; C=35 No-X=15; C=15 Omissions - X=55 C=57

d. Do you feel that you have been able to gain satisfactory proficiency in your normal job routines within a reasonable period of on-the-job training? ☐ yes, ☐ no. If no, explain.*

Yes-X=109; C=121 No-X=25; C=14 Omissions - X=3 C=2

e. What type maintenance would you prefer to perform in your unit? Check ☑ one block to indicate your first choice.

☐ Flight line X=21 C=18
☐ Combined flight line and shop X=100 C=93
☐ Periodic dock X=0 C=0
☐ No preference X=3 C=3
☐ Other (explain) Shop X=13 C=23

4. GENERAL:

a. Check ☑ one of the blocks below to indicate your opinion of the overall adequacy of training at Keesler in preparing you for your current job assignment.

☐ Inadequate	☐ Adequate	☐ More than Adequate
X=83	X=53	X=1
C=12	C=120	C=5

b. Indicate below any suggestions you may have for improving the technical training you received at Keesler.

Your Name
Rank
Unit
Date

* Use additional sheets if necessary.

BEST AVAILABLE COPY

Appendix B
Sample of Delayed Field Evaluation Questionnaire (Graduate)

BEST AVAILABLE COPY

N-X = 137 N-C = 137

1. HOW ESSENTIAL to your present job is a practical knowledge of, and the ability to apply successfully, maintenance principles related to each of the subject areas listed below?

a. NOT IMPORTANT: This is a "nice to know" subject area seldom, if ever, used by me in my current job.

b. SOMEWHAT IMPORTANT: My understanding of and ability to apply maintenance principles in this subject area are applicable to only a small portion of my current job.

c. HIGHLY IMPORTANT: I should understand and be able to apply principles in this subject area because it applies to a major portion of my current job assignment.

d. CRITICALLY IMPORTANT: My understanding of and ability to apply maintenance principles in this subject area are indispensable regardless of the frequency of its application in my current job.

2. ADEQUACY OF TRAINING: Indicate if training received at Keesler was inadequate, adequate or more than adequate.

INADEQUATE
ADEQUATE
MORE THAN ADEQUATE

SUBJECT AREAS		M	A	I	CRIT	HIGH	SOME	NOT
DC Fundamentals	X	0	34	103	35	63	34	3
	C	10	124	2	43	53	35	2
AC Fundamentals	X	0	27	110	38	63	32	1
	C	6	121	6	42	57	32	2
Vacuum Tubes	X	0	60	77	27	75	30	1
	C	8	118	8	32	75	22	4
Solid State Devices	X	0	11	126	26	43	53	12
	C	3	66	66	33	53	39	8
Basic Electronic Circuits	X	0	19	118	63	64	6	1
	C	7	115	13	57	51	24	0
RF Generation and Transmission	X	0	44	91	32	59	39	3
	C	3	111	21	35	67	28	2
Receivers	X	1	84	52	59	67	8	0
	C	8	121	6	70	57	7	0
Transmitters	X	0	81	55	[illegible]	63	9	0
	C	5	114	15	69	59	6	0
Timers	X	0	41	95	20	48	44	22
	C	6	89	40	18	46	43	26
Indicators	X	1	74	61	24	71	36	2
	C	4	93	39	26	64	36	7
Power Distribution and Control Circuits	X	2	60	74	25	54	34	1
	C	1	113	21	34	62	31	5
Test Equipment	X	24	101	12	69	56	8	0
	C	4	85	47	68	46	8	1
Troubleshooting Logic	X	19	101	17	88	41	4	0
	C	4	87	45	95	36	3	0
System Fault Isolation	X	20	91	26	86	44	2	0
	C	2	88	45	83	43	5	1
Electronic Circuit Tuning and Alignment	X	13	81	42	62	56	15	1
	C	3	97	36	59	61	12	0
Repair Techniques	X	7	98	31	49	73	9	0
	C	3	87	46	53	69	9	1
Use of Handtools	X	27	102	6	31	74	26	1
	C	12	98	26	28	74	28	3
Use of Maintenance Publications	X	43	83	11	46	65	12	0
	C	0	53	83	58	61	15	0
Spare Parts Acquisition	X	5	77	54	15	79	36	4
	C	0	63	73	25	63	39	6
Other All omissions not shown in totals.								

BEST AVAILABLE COPY

Appendix C

Sample of Delayed Field Evaluation (Supervisor)

N = X-141 N = C-142

1. HOW ESSENTIAL to his job requirements is the graduate's practical knowledge of, and ability to apply successfully, principles related to each major subject listed below.

a. NOT IMPORTANT: This is a "nice to know" subject area seldom, if ever, required by the graduate in his current job.

b. SOMEWHAT IMPORTANT: Graduate should understand and be able to apply principles in the subject area but it is only applicable to a small portion of his current job assignment.

c. HIGHLY IMPORTANT: Graduate should understand and be able to apply principles in the subject area because it is applicable to a major part of his current job assignment.

d. CRITICALLY IMPORTANT: Graduate's understanding and ability to apply principles in the subject area are indispensible regardless of the frequency of its application to his current field duties.

2. GRADUATE'S JOB PROFICIENCY: Indicate the graduate's demonstrated proficiency in applying the principles of each subject area below to the requirements of his job.

OUTSTANDING
SATISFACTORY
UNSATISFACTORY

SUBJECT AREAS		U	S	O	CRIT	HIGH	SOME	NOT
DC Fundamentals	X	35	96	7	63	65	11	0
	C	3	102	33	68	58	13	0
AC Fundamentals	X	40	92	7	62	66	10	0
	C	4	101	33	68	58	13	0
Vacuum Tubes	X	32	99	8	48	79	12	1
	C	2	107	29	47	77	14	0
Solid State Devices	X	66	68	5	45	53	36	5
	C	30	95	13	36	58	39	5
Basic Electronic Circuits	X	42	88	9	74	58	6	1
	C	5	99	34	77	52	9	0
RF Generation and Transmission	X	38	89	10	45	69	23	1
	C	11	103	23	43	73	21	1
Receivers	X	17	105	17	75	60	4	0
	C	3	100	35	72	61	6	0
Transmitters	X	19	104	16	71	68	5	0
	C	7	99	31	69	64	6	0
Timers	X	38	87	6	34	57	31	15
	C	13	106	13	32	53	41	12
Indicators	X	24	104	10	44	76	18	1
	C	5	103	19	40	80	16	2
Power Distribution and Control Circuits	X	25	103	10	56	69	12	1
	C	6	106	26	56	69	13	0
Test Equipment	X	10	106	23	92	43	4	0
	C	10	100	28	92	42	5	0
Troubleshooting Logic	X	23	87	29	104	34	1	0
	C	17	80	41	102	37	0	0
System Fault Isolation	X	21	94	23	96	40	2	0
	C	13	81	43	92	43	3	0
Electronic Circuit Tuning and Alignment	X	21	96	22	83	43	13	0
	C	8	96	34	77	48	14	0
Repair Techniques	X	13	96	30	83	53	3	0
	C	10	89	39	86	48	5	0
Use of Handtools	X	3	98	38	57	77	6	0
	C	3	100	35	55	80	4	0
Use of Maintenance Publications	X	11	94	34	66	63	10	0
	C	10	104	24	57	74	8	0
Spare Parts Acquisition	X	8	103	26	34	70	28	3
	C	5	111	21	36	73	27	2
Other (Identify) All omissions not shown in totals.								

Appendix C

Sample of Delayed Field Evaluation Graduate-on-the-Job Trainability Questionnaire (Unit Project Technician)

*N = X-116 *N = C-111

b. Check the appropriate squares in the table below to reflect your best estimate of the graduate's performance in the *TECHNICAL TRAINING* programs of paragraph 4a.

	X	C	X	C	X	C	X	C	X	C
GRADUATE TECHNICAL TRAINING PERFORMANCE										
Rating Scale	Far Below Average		Below Average		Average		Above Average		Far Above Average	
TRAINING TASKS										
Interest and Enthusiasm	0	3	8	4	33	29	45	36	13	19
Participation in Training Class Discussions	0	2	5	4	46	41	41	30	5	14
Ability to Ask Thoughtful Questions	0	1	5	5	47	32	42	35	3	17
Assistance and Counseling Required	2	9	22	14	53	40	18	16	2	10
Understanding of Technical Terminology	1	0	23	4	43	43	25	23	5	21
Understanding of Electronic Principles	2	0	33	2	32	40	26	30	4	18
Practical Application of Electronic Principles	2	1	23	5	41	38	27	28	4	18
Practical Knowledge of Basic Circuits	6	0	33	8	33	39	20	21	5	21
Application of Troubleshooting Logic	3	2	15	5	40	40	32	24	7	19
Use and Interpretation of Test Equipment	0	0	7	6	52	37	32	30	4	17
Use and Interpretation of Technical Data	0	0	8	6	52	41	31	27	6	17
Overall Trainability	0	0	9	6	49	34	32	31	7	28

* All training tasks were not rated by each Unit Project Technician.

8. In your judgment which of the following courses in the DOD/USAF Service Test Program would you recommend for training future radar repairmen?

X	C		
80	107	☐	"C" (Regular Control Course)
27	10	☐	"X" (Cue-Response Course)
15	6	☐	"Y" (Applied Electronics Course)

Briefly explain your reason for selecting the above course.

_____

_____

_____

9. Compared with other "5" skill level repairmen with whom you have worked, how do you rate yourself on the following evaluation items? (Check ☑ one appropriate square for each item.)

EVALUATION ITEMS	X	C	Far Below Average	X	C	Below Average	X	C	Average	X	C	Above Average	X	C	Far Above Average
Amount of Useful Work Produced	0	0		9	0		68	57		42	58		3	8	
Quality of Work Produced	0	0		4	0		67	43		45	61		6	19	
Speed in Completing Assigned Maintenance Tasks	1	1		18	3		64	66		34	39		5	14	
Ability to Work Without Close Supervision	0	0		5	2		57	39		53	59		7	23	
Understanding of Technical Terminology	10	0		31	2		61	73		18	40		2	8	
Understanding of Electronic Principles	29	0		67	11		22	68		2	35		2	9	
Practical Application of Electronic Principles	19	0		57	13		38	64		7	36		1	10	
Practical Knowledge of Basic Circuits	25	0		60	17		32	62		4	36		1	8	
Application of Trouble-shooting Logic	1	0		14	1		68	54		36	49		3	19	
Use and Interpretation of Test Equipment	1	1		6	9		82	65		29	39		4	9	
Use and Interpretation of Technical Data	1	1		12	6		79	77		27	33		3	6	
General Technical Ability	1	0		8	1		85	70		28	40		0	12	

Appendix D

Sample of Second-Job Proficiency Evaluation (Supervisor)

*N=X-134 C-135

1. SUPERVISOR'S EVALUATION OF GRADUATE JOB KNOWLEDGE

INSTRUCTIONS: Please check each knowledge item listed below in the appropriate column to indicate your judgment of the graduate's subject knowledge level. Base the rating on your observation of this graduate's performance while under your supervision.

SUBJECT KNOWLEDGE LEVELS	Scale Value	DEFINITION
	NA	Not applicable to or has not demonstrated his knowledge in his current assignment.
	D	EVALUATION. Can evaluate conditions and make decisions about the subject.
	C	ANALYSIS. Can analyze facts and principles and draw conclusions about the subject.
	B	PRINCIPLES. Can explain relationship of basic facts and state general principles about the subject.
	A	FACTS. Can identify basic facts and terms about the subject.

TRAINING STANDARD ITEM	X A C	X B C	X C C	X D C	X NA C
ELECTRONIC FUNDAMENTALS	/ / / /	/ / / /	/ / / /	/ / / /	/ / / /
1. D C Circuits	13 6	27 16	32 41	46 61	14 10
2. A C Circuits	13 7	34 19	27 41	44 57	14 10
3. Solid State Devices	29 14	31 31	21 34	18 30	33 25
4. Vacuum Tubes	12 5	30 16	32 47	44 56	14 10
5. Oscillators	21 8	26 24	43 46	27 45	15 11
6. Receivers Principles	15 5	26 22	41 48	34 43	16 16
7. Motors and Servomechanisms	16 5	28 24	42 44	26 41	20 19
8. Waveshaping Circuits	19 11	26 21	40 53	28 38	19 11
9. Microwave Principles	15 11	25 22	24 29	18 25	50 47
FUNCTION AND TACTICAL USE OF ACW RADAR EQUIPMENT	/ / / / / /	/ / / / / /	/ / / / / /	/ / / / / /	/ / / / / /
10. Search Radar	8 5	17 9	25 30	24 28	58 62
11. Height Finder Radar	4 6	15 10	25 24	48 42	40 52
12. Gap Filler Radar	8 6	7 3	5 8	7 5	105 112

BEST AVAILABLE COPY

N=X-134 C-135

Appendix D
Sample of Second-Job Proficiency Evaluation (Supervisor)

	Scale Value	DEFINITION
TASK PERFORMANCE LEVELS	NA	Not applicable to or not performed by the graduate in his current assignment.
	4	HIGHLY PROFICIENT. Can do the complete task quickly and accurately. Can tell or show others how to do the task.
	3	COMPETENT. Can do all parts of the task. Needs only a spot check of completed work. Meets minimum local demands for speed and accuracy.
	2	PARTIALLY PROFICIENT. Can do most parts of the task. Needs help only on hardest parts. May not meet local demands for speed or accuracy.
	1	EXTREMELY LIMITED. Can do simple parts of the task. Needs to be told or shown how to do most of the task.

TRAINING STANDARD ITEM	X 1	C	X 2	C	X 3	C	X 4	C	X NA	C
10. Uses maintenance data collecting forms	1	1	12	18	54	51	44	49	23	16
ADJUSTS AND USES TEST EQUIPMENT SUCH AS	/	/	/	/	/	/	/	/	/	/
11. Multimeters	0	0	9	0	42	42	71	84	12	9
12. Vacuum tube voltmeters	0	0	11	1	42	45	68	79	13	10
13. Signal generators	3	2	16	9	60	63	40	53	15	8
14. Oscilloscopes	4	1	10	6	54	54	54	65	12	9
15. Spectrum analyzers	9	5	27	23	46	57	23	28	29	22
16. Power meters	5	4	23	10	39	53	48	54	19	14
17. Crystal checkers	6	1	18	17	36	49	22	22	52	46
19. Transistor testers	8	2	19	22	40	37	9	19	58	55
20. Frequency measuring devices	8	2	23	19	49	48	27	48	27	18
21. MTI evaluators	9	3	11	11	23	19	4	13	87	89
22. Transponder test sets	6	3	9	12	19	30	23	15	77	75
23. Selects the proper test equipment and repair tools in maintaining ACW equipment	3	1	15	9	56	58	48	58	12	9
24. Recognizes defective test equipment through indications obtained during operation	5	3	12	15	56	49	46	59	15	9

Appendix E
Sample of First Enlistment Survey (Graduate)

2. Date separated from the Air Force:

_____ day _____ month _____ year.

3. Base of separation:

_____

4. Military pay grade at the time of separation from the Air Force:

☐E-1, ☐E-2, ☐E-3, ☐E-4, ☐E-5, ☐Other_____

5. Marital status:

a.☐ Single b.☐ Married c.☐ Divorced d.☐ Widower

6. Number of dependents:

☐1, ☐2, ☐3, ☐4, Other_____

7. Do you have any plans at present for returning to active military duty? ☐ Yes, ☐No. If yes, please list date. ____ day ____ month _____ year.

8. Are you now a member of an active reserve unit? ☐Yes, ☐No.

9. In your judgment, did the technical training you received at Keesler AFB adequately prepare you for satisfactory career progression in the Air Force? ☐ Yes, ☐ No. If no, please explain briefly. _____

_____

10. Please check ☑ the following statement concerning your employment:

☐Full Time, ☐Part Time, ☐Unemployed

11. If you are now self-employed, give a brief description of your work. _____

_____

_____

12. If you are now employed by someone other than yourself, please complete the following items:

a.☐ Name of employer _____

b.☐ Type business _____

c.☐ Brief description of work _____

_____

_____

13. If you are employed, does your job require any knowledge or skills on which technical training was provided at Keesler AFB? ☐Yes, ☐No. If yes, please list and explain briefly. _____

_____

_____

Appendix E

Sample of First Enlistment Termination Survey (Supervisor)

N = X-83 C-89

Observation: Is your rating based upon a personal observation of the graduate performing this task?

DEGREES OF SUPERVISION

4 Unsupervised: Can perform the whole task by himself without supervision or technical assistance. Can tell or show others how to perform the task.

3 Occasionally Supervised: Can perform all parts of the task. Requires only a spot check of completed work.

2 Frequently Supervised: Requires supervision/technical assistance practically throughout the task. Can perform only simple parts of the task unsupervised.

1 Continually Supervised: Requires supervision/technical assistance throughout the task.

PERFORMANCE TASKS	X^1C	X^2C	X^3C	X^4C	Yes	No
2. Uses of technical publications such as:	/ / /	/ / /	/ / /	/ / /	/ / / /	
a. Wiring diagrams	2 0	7 7	30 25	43 53		
b. Circuit diagrams	2 0	9 6	28 24	43 55		
c. Parts provisioning and illustrated parts breakdown	1 0	3 3	20 21	58 61		
d. Alignment procedures	2 0	7 4	26 27	44 51		
3. Troubleshooting of electronic equipment:	/ / /	/ / /	/ / /	/ / /	/ / / /	
a. Localizes malfunction to major assembly	1 0	4 1	16 21	59 63		
b. Localizes malfunction to subassembly	1 0	7 2	23 23	49 57		
c. Localizes malfunction to a circuit	2 0	14 7	26 36	38 39		
d. Localizes malfunction to a circuit component (such as capacitor, resistor, etc.)	3 0	19 6	25 40	31 35		
e. Restores equipment to performance standards	1 0	11 3	31 31	35 48		
4. Transfer of knowledge and skills from familiar to unfamiliar electronic equipment	3 1	19 11	38 39	19 26		

NOTE: All supervisors did not rate all graduates on all items.

PROCESS VERSUS PRODUCT MEASURES IN PERFORMANCE TESTING

William C. Osborn
Human Resources Research Organization

Consider the following situations.

After having undergone training in tank gunnery a soldier's proficiency is being tested on a gunnery range. During the course of this test he will fire several main gun rounds at targets varying in size, shape and distance. In each case his score is determined by whether he hit the target within some specified time limit, and he is certified a tank gunner if he scores above some minimum level required for qualification.

Under other circumstances, a soldier having undergone similar training may be evaluated differently. Let's assume that ammunition is scarce or that adequate range facilities are not available. Here the soldier might have to be tested under dry-firing conditions. He would be required to take actual or miniaturized versions of targets under fire, and a tester would assess in each case whether the gunner 1) acquired the target with smooth manipulation of the hand controller, 2) correctly ranged on the target, 3) achieved the proper sight picture, 4) squeezed the firing switch without losing the sight picture, and 5) fired within some allotted time. Here the gunner is qualified if he performed each of the five procedural steps correctly on some minimum number of targets.

In the first situation described, a task outcome or <u>product</u> measure -- target hits -- is the basis for evaluating gunners; whereas, in the second instance correct task procedure or a <u>process</u> measure is the basis for evaluation. Though somewhat oversimplified, the contrasting approaches to

performance testing drawn in these two examples illustrate the focus of this paper: the use of process versus product measures in performance testing.

I am chiefly interested in the use of performance tests to evaluate the results of training, and to properly set the stage for what I have to say today let me first summarize what the training evaluator considers to be the ideal use of product and process measures. Performance tests are used in training evaluation to serve two purposes: (a) to certify student achievement, and (b) to diagnose weaknesses in the instructional system. In the use of such tests, proficiency measures which focus on task outcomes (products) normally provide data relevant to the first purpose, whereas measures of how the tasks are carried out (process) pertain to the second. For example, the number of targets hit by the tanker trainee would be the product measure by which his qualification as a tank gunner is assessed. However, if he fails to qualify we would also like to know why -- where was his training weak? This is where process measures are useful: if the gunner consistently missed targets, was it because he ranged incorrectly, or obtained an improper sight picture, or wasn't able to maintain the gun lay during firing? This type of data is useful in diagnosing areas of training deficiency, and is essential in efficiently remediating trainees.

Thus we see the roles played by product and process measures in training evaluation. Both are important -- even critical -- when used for their respective purposes in evaluating the results of training.

Product and Process as Measures of Student Achievement

With this background I would now like to narrow the focus of my comments to the use of these types of performance measures in serving just the first

of the two evaluation purposes stated above -- that of certifying student achievement. In testing a student to determine if he is qualified to advance to the next level of training, or ultimately out of training and on to the job, we would, as mentioned, normally prefer to use a product score. Before a man is certified as a gunner we would like to have him demonstrate that he can hit targets; or before certification as a navigator he should actually demonstrate that he can get from point A to point B; etc.

Although it may safely be said that every task has a purpose -- the fact of the matter is that in practice a great many performance tests are used which employ process measures only in evaluating student achievement or job readiness. Why is this the case? Is the substitution of process for product measurement justified? If so, when? If not, how may the test developer improve his methods? These are the questions that I will be addressing today.

Product, Process and Types of Tasks

Before exploring in more detail the issue of why process measures are so widely substituted for measures of task product, it will first be helpful to consider three categories of tasks:

1. Tasks in which the product is the process.
2. Tasks in which the product always follows from the process.
3. Tasks in which the product may follow from the process.

Relatively few tasks are of the first type, those in which product and process are one and the same. These are normally tasks which serve an aesthetic purpose such as gymnastic exercises or springboard diving. Close

order drill is a good military example. Here we see that the outcome or product of the task is no more or less than the correct execution of steps in task performance -- that is, the process.

More tasks are of the second type mentioned, those in which the product invariably follows from the process. Fixed-procedure tasks typically fall in this category. Troubleshooting an electrical circuit, disassembling a rifle, and implanting a land mine are examples. In tasks of this type the procedural steps are known, observable, and comprise the necessary and sufficient conditions for task outcome; so if process is correctly executed, task product necessarily follows.

A great many job tasks are of the third type where the product is less than fully conditional on the process. In other words, with these types of tasks the process may appear to have been correctly carried out but the goal or product not achieved. This can happen for one of two reasons: either (a) because we are unable to fully specify the necessary and sufficient steps in task performance or (b) because we do not or cannot accurately measure them. In aim-firing a rifle, for example, we are interested in knowing if a soldier is standing with his body properly oriented to the target, face properly positioned on stock, rifle sling in correct position, lead arm perpendicular, and firing arm parallel to ground; if he is breathing correctly, has a good sight picture, and squeezes the trigger. Presumably, if this process is followed the rifleman will hit the target. Assuming that we have identified all essential steps in rifle firing, and, further, that we can reliably measure their correct execution, then the task is of the second type described above and does not belong in Category 3. However, in practice, we know that our best efforts to evaluate execution of this particular task are not sufficient to warrant substituting process for product measurement.

In other words, sometimes the target is missed even though in the judgment of a skilled evaluator the rifleman did everything right. Therefore, either because we are not absolutely certain that we have identified all necessary steps in the firing process or because we cannot accurately assess the execution of some of them we ultimately qualify a rifleman on the basis of whether he hits the target.

In reflecting on the nature of these three types of tasks an important implication emerges regarding the role of product measurement in testing task performance: Because of the interchangeability of process and products for tasks of the first two types, it doesn't really matter which measure is used to assess proficiency; but for tasks of Type 3, product measurement is very important. In spite of this, in practice, performance tests for many of the latter type of tasks do not attempt to measure product. Why is this so?

Problems in Product Measurement

The reasons largely stem from practical considerations in which the measurement of task product is viewed as either too costly, too dangerous, or for other reasons simply too impractical. In testing such performances as hand-to-hand combat, for example, where task product would take the form of disabling a hostile enemy, the test developer is normally limited to requiring the demonstration of task process. Similarly, in a first aid task like controlling the bleeding from an external wound, the person being tested is, for obvious reasons, asked only to demonstrate the process. Or, in removing a jammed round from a weapon, it is considered impractical to actually jam a round in order to create a valid test situation, so again only the steps in task performance are measured. Many such examples are to

be found in the area of interpersonal behavior. One has to do with instructor training, where at least traditionally the military instructor trainee is evaluated by having him prepare and deliver a block of instruction during which he is judged on such process factors as: "stood erect," "had good eye contact with audience," "could be heard in the back of the room," "used visual aids effectively," "covered all points in the lesson plan," etc. Although the product of instruction clearly is student learning, I believe it is seldom if ever used as the criterion for qualifying an instructor trainee -- probably because it would involve a more time consuming and impractical method of evaluation. Another very similar example which comes to mind pertains to a recruiter's task of delivering a persuasive speech to a student audience. If the product of this task could be measured it would be in terms of the number in the audience who later contact the recruiter with an interest in enlisting. But, again, because of its implausability as a measure of student achievement, product gives way to process and the recruiter trainee's persuasive speech is evaluated in much the same way as was described above for instructor trainee.

Dealing with Problems of Product Measurement

I'm sure that those of you involved in performance testing can think of many more instances in which product measurement is not used. Clearly, some of these are justified by cost or safety considerations -- but others are not. Which brings us to the central point of my remarks today. I believe that test developers often fail to see the importance of measuring task outcome; or perhaps they merely slight the importance when faced with practical limitations in its measurement. Whatever the motivation, I believe they do not strive nearly hard enough to overcome resource problems which

constrain attempts to measure task product, and too easily give in to the simplistic approach of measuring task process. The overriding question that a test designer should ask himself in this situation is: If I use only a process measure to test a man's achievement on a task, how certain can I be from this process score that he would also be able to affect the product or outcome of the task? Where the degree of certainty is substantially less than that to be expected from normal measurement error, the test designer should pause and reconsider ways in which time and resource limitations can be compromised in achieveing at least an approximation to product measurement. Although there will remain instances in which product measurement simply cannot be achieved, we will discover many others where, through some imaginative thinking, we can devise simulations that will enable us to assess task outcome in a more relevant fashion.

In testing the student instructor, for instance, I see no compelling reason why we shouldn't get away from the "charm school" approach to evaluation. Why not simply have him conduct a brief instructional session for a small group of students (perhaps his peers), with his achievement being measured in terms of whether his students have accomplished the instructional objective? In the case of the recruiter trainee's speech, evaluating task product is more difficult; but surely a measure closer to task outcome could be achieved -- perhaps a paid student panel representing the potential audience could be employed to view and rate the appeal of videotaped trainee speeches. In zeroing in on critical motor skills, such as those involved in extracting a jammed round from a weapon or in controlling bleeding from a wound, it would seem that relatively low cost simulators could be devised for use in testing task outcome. Hand-to-hand combat very likely represents a case in which ultimate task product simply cannot be measured. However,

in a similar vein, the Army is now experimenting with an intriguing method of assessing the outcome of an infantry squad combat exercise. The principal feature of the method entails each participant having a number printed on his helmet and an inexpensive scope mounted on his rifle; then, during the course of the exercise a soldier may "kill" by correctly reporting an enemy's number, or "be killed" by allowing his number to be sighted by the enemy. Number size and scope power have been carefully calibrated from empirical data so that the probability of a simulated "kill" is highly correlated with the expected outcome in actual battle. This is an excellent example of an innovative method of achieving product measurement on a task that heretofore had been subject to process evaluation.

Obviously, from these examples we can see that the accomplishment of product measurement is not always a simple matter; but it is a demanding and essential goal to be pursued by the performance test developer if his products are to be relevant to real world behavior.

BEST AVAILABLE COPY

BRIEF OF FOUR RESEARCH STUDIES USING THE SELF-EVALUATION TECHNIQUE (SET STUDIES)

US Army Ordnance Center and School
Doctrine and Training Development Division

Purpose of the SET Studies

The purpose of the four SET studies was to determine if student self-evaluations of their performance tests improved student performances on required performance tests.

Procedure of the SET Studies

This section describes the subjects, course requirements, testing instruments used, how the tests were administered, and how the scores were obtained.

The Subjects

The subjects used in these SET studies numbered three hundred and fifty-four students who performed twenty-seven hundred and four performance tests. These students were enrolled in the 44C20 Welder Course, 41C10/20 Fire Control Instrument Repairman Course, 44E20 Machinist Course, and 63G20 Fuel and Electric Systems Repairman Course. These courses were conducted at the US Army Ordnance Center and School.

Course Requirements

One of the prerequisites for each of the courses is as follows: Welder, General Maintenance (GM) score of 90, Fire Control Instrument Repairman, General Maintenance (GM) score of 100, Machinist, General Maintenance (GM) score of 100, and Fuel and Electric Systems Repairman, Motor Maintenance (MM) score of 95. These Army Standard Scores are derived from the General Maintenance and Motor Maintenance Aptitude Areas of the Army Classification Battery.

The Instruments Used in the Studies

For the evaluation of the apprentice welders one evaluation form was designed and served the purpose of evaluating the eight performance tests. Any one of the eight welds performed can be written in the blank space on the instrument. A copy of this instrument is attached as Appendix A.

For the evaluation of the Fuel and Electric Systems Repairman it was necessary to break down several of the performance tests so that some of the required tasks could be tested at a prescribed testing station. A copy of this instrument is attached as Appendix B.

For the evaluation of the Fire Control Instrument Repairman and the Apprentice Machinist a separate performance evaluation instrument was designed to cover each of the performance tests. Copies of these instruments are attached as Appendix C and D.

Testing Procedure

Several classes were used to validate all self-evaluation instruments. After the instruments were validated, students who did not make self-evaluations were designated as the control group and students making self-evaluations were considered to be the experimental group.

The SET test project director matched each student in the control group with a comparable student in the experimental group using the GM or MM scores as a basis for equating these matched pairs. In some instances it was impossible to make comparable matches for each of the matched pairs when matching the students by classes. The overall results on the total of matches for each course showed a very close mean for the matched pairs.

The experimental groups made a self-evaluation for each of the performance tests they performed. A technically qualified grader evaluated each performance test performed by both the experimental and control groups. These evaluations were forwarded to the SET study project director who scored the self-evaluator's score sheets, the grader's evaluation of the self-evaluators' score sheets and the graders' evaluations of the non-self-evaluations. The student's self-evaluation and the grader's evaluation sheets were attached together and returned to the students with either critical or complimentary remarks which were influenced by the scores they made on that performance test. No score was recorded on the returned student or grader's evaluation sheets and the comments reflected only his weak or strong tasks on the performance test. The graders' scores for the non-self-evaluators were recorded by the project director opposite his match from the self-evaluators.

The performance test scores for each of the matched pairs were used to test for a significant difference between the means of the control and experimental groups.

For the purpose of evaluation the SET studies were designed to test for significant difference between the graders' scores for students who made self-evaluations and the graders' scores for students who did not make self-evaluations.

BEST AVAILABLE COPY

The secret of the set studies which I am going to present can be attributed to direct communication and automatic feedback.

To show you the importance of proper and correct communicative procedure I would like to read a copy of the registration form I filled out when I registered for this military testing association symposium.

Now I will briefly cover the need for conducting set studies. John B. Carrol's article "Neglected Areas in Educational Research" presented in number 42, May 1961 issue of the Phi Delta Kappan stated that "Research has told us little about the role of motivation in school learning." "Let us take for granted that motivation is the sense of willingness on the part of the learner to engage in learning." Studies of both intrinsic and extrinsic motivation are needed.

S. L. Pressey and F. P. Robinson stated in their article "Psycholog and the New Education." Informing students regarding their work is essentia for intelligent learning. If real life tests are brought to the point where both student and teacher can be informed about progress in the development of social adjustments, interests and attitudes this would be splended.

The only study which I could find that even vaguely resembled the SET studies was located in the twenty-ninth yearbook of the National Society for the study of education and was published in 1930.

THE DESIGN AND RESULTS OF THE STUDY ARE AS FOLLOWS:

Two equivalent groups of 358 Fourth grade pupils were given identical arithmetic drill exercises for 15 minutes a week for twenty-one weeks. The only variable was specific knowledge of improvement. The members of the control group never received their weekly scores. Those in the experimental classes kept individual progress charts and pooled their losses or gains in a graphic representation of improvement of the class as a whole. The record of actual improvement was made possible by using drill units for which comparable standards had been provided. This scoring device compensated for difference in difficulty of the tasks. The group which had continuous information concerning improvement made significantly greater gains than the control group.

This indicates that if students are knowledgable of what requirements are expected of them and they are made aware of their progress there should be an improvement in their overall performance.

Please refer to the self evaluation instrument for the welders. (Appendix A)

Several classes were used to validate this instrument. The first, second and third instruments were discarded and this instrument was

adopted for the evaluation. This same instrument can be used for all eight welding projects. All that is needed is to unite the type of weld being performed in the lined space provided on the instrument.

Are there any questions?

The same identical procedure was followed for the next 3 SET studies.

Now refer to the second instrument Appendix B which is a sample of one of the performance tests given to the 63G20 Fuel and Electric Systems Repairman course students. The development of instruments to evaluate the 33 matched pairs for 7 performance tests was complicated by the number of stations needed to accomplish the evaluations. For the seven tests administered a total of 22 sub tests were needed to cover the 22 stations covered by the 7 performance tests. It was also difficult to set up required standards for each of the tasks required to be performed. Notice that this was solved by the reference to instructor help. The involvement of the instructor determines how well the student could perform each required task.

Are there any questions?

Please refer to the instrument for the evaluation of the 41C instrument repairman, Appendix C. Eight performance tests were developed with a separate test instrument developed for each performance test. The sample which has been furnished you is an example of the task requirements for performance test number. Six of eight tests were easier to construct because the students had to perform tasks to meet required standards.

Are there any questions?

The last sample evaluation instrument is for the machinist course and was written to cover the first performance test, Appendix D. The remaining 6 performance tests were written following the same format, and no problem was encountered since each of the tests had tolerance ranges which were indicative of the type of performance the student made.

Do I have any questions?

Grading of instruments

The grading of the instruments is very simple. You simply take the number of tasks and multiply by 4. For example on the last instrument you have 10 tasks x value of 4 equal score of 40. Suppose the student had 6 tasks in column 4 = 24; 3 tasks in column 3 = 9; and 1 task in column 2 = 2 for a total of 35; dividing 4 into 35 would give a grade of 87.5.

BEST AVAILABLE COPY

Please refer to Appendix E which shows the final results for the four studies. It shows that the self evaluation technique was successful for all the studies conducted. A copy of the complete study can be obtained upon request by writing:

Dr. John J. Holden
US Army Ordnance Center & School
ATTN: ATSL-CTD-DT-P
Aberdeen Proving Ground, Maryland 21005

BEST AVAILABLE COPY

APPENDIX A

EVALUATION OF STUDENT'S
WELDING PROJECT

STUDENT'S NAME _____

CLASS NO _____

DATE _____

MARK AN (X) IN THE PROPER SPACE

STUDENT SELF-EVALUATION _____

GRADER EVALUATION _____

INTRODUCTION: This is a complete welding project which will indicate how well you feel you performed the important tasks for a _____ _____weld.

The purpose of this experiment is to obtain information that will help us to improve the welding course. You will not receive a grade on your welding project but an honest evaluation of your performance will be of a great help to us.

DIRECTIONS: Read each item on the task list and mark an (X) in the column which shows how well you feel you performed that task.

	Column 4	Column 3	Column 2	Column 1
TASKS PERFORMED	Equal to Performance of Better than Average Student Welder	Equal to Performance of Average Student Welder	Not quite Equal to Performance of Average Student Welder	Far Below the Performance of Average Student Welder
1. Obtained correct flame adjustment (current setting).				
2. Obtained good fusion (tinning action).				
3. Obtained good crown on weld face.				
4. Obtained good penetration (bonding).				
5. Kept bead free of cracks.				
6. Obtained even uniform bead shape.				
7. Kept bead free of undercut (droplets).				
8. Kept bead free of overlap.				
9. Kept bead free of burned or crystallized metal.				
10. Kept bead free of holes.				

Could finish project in time allotted YES ____ NO____

APPENDIX B

PERFORMANCE TEST 63G20

DELIVERY VALVE SPRING TEST

STATION 1 Checking Delivery Valve

Student's Name_____ ☐ Student Self-Evaluation

Class #_____Date_____ ☐ Instructor Evaluation

1. Connected the nozzle tester and prepared the pump for testing the delivery valve springs with:

   ☐ No trouble ☐ Little trouble ☐ Lots of trouble

   Received instructor's help to perform operation: ☐ YES ☐ NO

   If YES is checked, how many times did the instructor help? ☐

2. Observed opening pressure of all delivery valve springs, recorded readings and determined condition of valves:

   ☐ 6 valves correctly ☐ 4 or 5 valves correctly ☐ 3 or less valves correctly

   Received instructor's help to perform operation: ☐ YES ☐ NO
   If YES is checked, how many times did the instructor help? ☐

3. Overall performance for this station:

   ☐ Above average student performance ☐ Equal to average student performance ☐ Below average student performance ☐ Far below average student performance

BEST AVAILABLE COPY

BEST AVAILABLE COPY

APPENDIX B (Cont)

PERFORMANCE TEST 63G20

STATION 2 Timing APE6BB Injector Pump to the Mack ENDT-673 Engine

Student's Name_____ ☐ Student Self-Evaluation

Class #_____Date_____ ☐ Instructor Evaluation

1. Positioned number 1 cylinder in correct position on compression stroke with timing mark aligned to:

☐ 29° BTDC ☐ 38° or 30° BTDC ☐ 27° or 31° BTDC

Received instructor's help to perform operation: ☐ YES ☐ NO
If YES is checked, how many times did the instructor help? ☐

2. Secured pump drive coupler to engine with pump camshaft at "part closing" position for number 1 outlet with:

☐ No trouble ☐ Little trouble ☐ Lots of trouble

Received instructor's help to perform operation: ☐ YES ☐ NO
If YES is checked, how many times did the instructor help? ☐

3. Replaced all lines, covers, and valves and engine started with:

☐ No trouble ☐ Little trouble ☐ Didn't start

Received instructor's help to perform operation: ☐ YES ☐ NO
If YES is checked, how many times did the instructor help? ☐

4. Overall performance for this station:

☐ Above average student performance ☐ Equal to average student performance ☐ Below average student performance ☐ Far below average student performance

BEST AVAILABLE COPY

APPENDIX B (Cont)

PERFORMANCE TEST 63G20

STATION 3 Timing the Unit Injectors

Student's Name_____ ☐ Student Self-Evaluation

Class #_____Date_____ ☐ Instructor Evaluation

1. Positioned engine for timing of the selected injector correctly:

☐ No trouble ☐ Little trouble ☐ Lots of trouble

Received instructor's help to perform operation: ☐ YES ☐ NO
If YES is checked, how many times did the instructor help? ☐

2. Selected correct timing gage and adjusted injector to specifications with:

☐ No trouble ☐ Little trouble ☐ Lots of trouble

Received instructor's help to perform operation: ☐ YES ☐ NO
If YES is checked, how many times did the instructor help? ☐

3. Overall performance for this station:

☐ Above average student performance ☐ Equal to average student performance ☐ Below average student performance ☐ Far below average student performance

Did you have enough time to complete the test required at each of the 3 stations?
☐ YES ☐ NO. If answer is NO, write reason(s) on back of this sheet.

APPENDIX C

SIXTH PERFORMANCE EXAMINATION 41C10/20

Maintenance and Adjustments of Panoramic Telescope M115

STUDENT'S NAME _____

CLASS NUMBER _____

DATE _____

MARK AN (X) IN THE PROPER BOX

STUDENT SELF-EVALUATION ☐

GRADER EVALUATION ☐

PURPOSE: The purpose of this self-evaluation experiment is to obtain information that will help us to improve the fire control instrument repair course. You will not receive a grade on your self-evaluation for this project but an honest evaluation of your performance will be of a great help to us to evaluate the course. Your course grade will be determined by the score you obtain from the department's performance examination.

DIRECTIONS: Mark a ✓ in the TOLERANCE RANGE COLUMN for the TOLERANCE which you obtained on each of the listed dimensions to be checked.

TOLERANCE RANGE

DIMENSIONS TO BE CHECKED	Column 4	Column 3	Column 2	Column 1
1. Parallelism of reticle and FOV	Perfect ☐	Within ½ Mil ☐	Within ½ Mil ☐	Over ½ Mil ☐
2. Magnification of FOV	Perfect ☐	Over-Under 1/8 Mil ☐	Over-Under ¼ Mil ☐	Over-Under ½ Mil ☐
3. Collimation	Perfect ☐	Within ¼ Mil ☐	Within ½ Mil ☐	Over ½ Mil ☐
4. Parallax	No Parallax ☐	¼ Mil ☐	½ Mil ☐	Over ½ Mil ☐
5. Eyepiece Focus	Perfect ☐	Within ¼ Diopter ☐	Within ½ Diopter ☐	Over ½ Diopter ☐
6. Adjustment of Counter Mechanism	Within ½ Mil ☐	Within ½ Mil ☐	Within 3/4 Mil ☐	Over 3/4 Mil ☐
7. Backlash	Perfect ☐	Within ¼ Mil ☐	Within ½ Mil ☐	Over ½ Mil ☐
8. Sideplay	☐	☐	☐	☐
9. Lift	☐	☐	☐	☐
10. Cleanliness of Optics	Very Good ☐	Good ☐	Fair ☐	Poor ☐
11. Instructor Assistance	No Assistance ☐	Assistance 1 Time ☐	Assistance 2 Times ☐	Assistance More Than 2 Times ☐
12. Evaluation of Completed Pan Tel	Above Average ☐	Average ☐	Below Average ☐	Far Below Average ☐

APPENDIX D
FIRST PERFORMANCE EXAMINATION
EVALUATION OF STUDENT'S
MACHINIST PROJECT
SHAPER PERFORMANCE TEST

STUDENT'S NAME_____

CLASS NUMBER_____

DATE_____

MARK AN (X) IN THE PROPER BOX

STUDENT SELF-EVALUATION ☐

GRADER EVALUATION ☐

PURPOSE: The purpose of this self-evaluation experiment is to obtain information that will help us to improve the machinist course. You will not receive a grade on your self-evaluation for this project but an honest evaluation of your performance will be of a great help to us to evaluate the course.

DIRECTIONS: Make a √ in the TOLERANCE RANGE COLUMN for the TOLERANCE which you obtained on each of the listed dimensions to be checked.

DIMENSIONS TO BE CHECKED	DIMENSION	Column 4 ± 0.010"	Column 3 ± 0.012"	Column 2 ± 0.014"	Column 1 Above or Below ± 0.014"
1. Height	1.750"	☐	☐	☐	☐
2. Width	1.750"	☐	☐	☐	☐
3. Depth from Top of V to Slot	0.625"	☐	☐	☐	☐
4. Width of First Step	0.625"	☐	☐	☐	☐
5. Width of Second Step	0.625"	☐	☐	☐	☐
		± 1/64"	± 1/32"	± 3/64"	Above or Below ± 3/64"
6. Length of Block	2 - 3/4"	☐	☐	☐	☐
7. Linear Pitch	25/64"	☐	☐	☐	☐
8. Width of V	1 - 3/16"	☐	☐	☐	☐
9. Depth of Teeth	0.270"	+ 0.005 - 0.000	+ 0.006 - 0.001	+ 0.007 - 0.002	Above + 0.007 Below - 0.002
		☐	☐	☐	☐
		Better Than Average	Average	Below Average	Far Below Average
10. General Appearance		☐	☐	☐	☐

Could you finish project in time allotted YES☐ NO☐. If answer is NO tell why you didn't have enough time. Did you finish ahead of time YES☐.

AD-A040 773 AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TEX F/G 5/9
PROCEEDINGS, ANNUAL CONFERENCE OF THE MILITARY TESTING ASSOCIAT--ETC(U)
NOV 73

UNCLASSIFIED NL

4 OF 9
ADA
040773



BEST AVAILABLE COPY

APPENDIX E

TABULATIONS OF SET STUDY FINDINGS
FOR 177 MATCHED PAIRS OF STUDENTS
WHO COMPLETED 2704 PERFORMANCE TESTS

44C20 Welder Course					
No. Matched Pairs 81	8 Performance Tests		Total No. Performance Test 1296		
GM Mean NSE	GM Mean SE	Significant Difference	PT Mean NSE	PT Mean SE	Significant Difference SE on 7 of 8 Tests
109.21	109.16	None	72.48	75.66	.01 level SE

63C20 Fuel and Electric Repairman					
No. Matched Pairs 33	7 Performance Tests		Total No. Performance Tests 462		
GM Mean NSE	GM Mean SE	Significant Difference	PT Mean NSE	PT Mean SE	Significant Difference
109.5	108.1	None	89.6	91.9	.01 level SE

41C 10/20 Fire Control Instrument Repairman					
No. Matched Pairs 32	8 Performance Tests		Total No. Performance Tests 512		
GM Mean NSE	GM Mean SE	Significant Difference	PT Mean NSE	PT Mean SE	Significant Difference
113.21	107.03	.01-NSE	83.6	86.2	.05 level SE

44E20 Machinist Course					
No. Matched Pairs 31	7 Performance Tests		Total No. Performance Tests 434		
GM Mean NSE	GM Mean SE	Significant Difference	PT Mean NSE	PT Mean SE	Significant Difference
114.4	109.4	.01-NSE	89.1	90.4	None SE

A MODULAR APPROACH TO PROFICIENCY TESTING

by

Robert W. Stephenson, Ph.D.
Warren P. Davis, Col. USA (Ret.)
Harry I. Hadley

AMERICAN INSTITUTES FOR RESEARCH

and

Mrs. Bertha H. Cory

U.S. ARMY RESEARCH INSTITUTE FOR THE
BEHAVIORAL AND SOCIAL SCIENCES

September 1973

From a paper presented at the Fifteenth Annual
Conference of the Military Testing Association,
San Antonio, Texas, 28 October to 2 November 1973

INTRODUCTION

A new, more specific language for describing work activities is being designed for the Army on an experimental basis.[1] The new language is based upon a concept called the "duty module". Duty modules are clusters of tasks that tend to go together occupationally and organizationally in meaningful ways.

The need to evaluate the feasibility of personnel information systems based upon clusters of tasks smaller than a Military Occupational Specialty (MOS) was originally suggested by personnel at the Army Research Institute. The development of the duty module concept was a team effort involving staff members from the Army Research Institute (ARI) and the American Institutes for Research (AIR).[2]

Duty modules are rationally derived clusters of work activity based upon a detailed examination and grouping of task inventory or job analysis data. Attention is then given to ways in which these tentatively identified job content modules can be tested against various available criteria of operational utility. One relevant consideration is whether job content modules can be used as field assignment modules. Another possible application is in the area of requirement planning and unit effectiveness. A source of information here is data that can be gathered in conjunction with unit training and unit effectiveness exercises that are performed in the field.

The word "module" was chosen because job activity clusters, like equipment components of the same name, are meant to be largely self-contained, independent units of work. For purposes of occupational classification, a duty module is a cluster of tasks that apply without modification in a number of occupational classifications or specialties.

---

[1] The work was carried out under Contract No. DAHC-19-71-C-0004, "A Taxonomic Base for Future Information and Decision Systems", and Contract No. DAHC-19-73-C-0041, "A Comparison of Officer Job Content Modules with Activity Grouping Implicit in Course Design", awarded by the Army Research Institute to the American Institutes for Research.

[2] Key personnel at ARI were Mr. Cecil Johnson, who provided the initial guidance as the Contract Officer's Technical Representative, Dr. J.E. Uhlaner, Technical Director, and Mrs. Bertha Cory, who succeeded Mr. Johnson as the Technical Representative.

Key AIR staff members, in addition to Dr. Robert W. Stephenson, who was Project Director, included Dr. Robert Miller, who served as Consulting Scientist, Colonel Warren P. Davis, Mr. Harry I. Hadley, Dr. Edwin A. Fleishman, Dr. Albert S. Glickman, Mr. Clifford P. Hahn, Dr. Ronald P. Carver, and Mr. Albert Farina.

Another form of construct validation is available through examination of mission elements which are activities that are used to design and evaluate the performance of organizational unit.

This presentation will describe two different kinds of modular evaluation devices -- sets of tasks performed by individuals, and sets of tasks performed by organizational units.[3]

INDIVIDUAL PROFICIENCY TESTS

Every individual proficiency test and every Army training test is already divided into special component sections with separate scores. Before going into detail about what modular component scores are supposed to do and supposed to look like, it is necessary to describe what these existing systems are like.

ENLISTED PERSONNEL PERFORMANCE EVALUATION

This discussion should be prefaced by noting that evaluation of the performance of enlisted personnel is an important responsibility of every commissioned and senior noncommissioned officer in the Army. The rewards and punishments associated with such evaluations give commissioned and noncommissioned officers the necessary control over enlisted personnel to maintain and improve effectiveness. In addition to this important supervisory function, however, there are a number of formal proficiency evaluation procedures for enlisted personnel that are conducted by various headquarters. The most important of these, for purposes of this paper, is the U.S. Army Enlisted Evaluation Center, located at Fort Benjamin Harrison, Indianapolis, Indiana.

---

[3] These developments are described in detail in Miller, R.B. A Taxonomic Base for Future Management Information and Decision Systems: Theoretical Background to the Design of Duty Modules; American Institutes for Research, Washington, D.C., Technical Report AIR-23500-7/71-TR-2, July 1971. (U.S. Army Behavior and Systems Research Laboratory, BESRL Technical Research Note, in preparation.); and Stephenson, R.W. (American Institutes for Research, Washington, D.C.) A Taxonomic Base for Future Management Information and Decision Systems: A Common Language for Resource and Requirement Planning; U.S. Army Behavior and Systems Research Laboratory, Arlington, Va., Technical Research Note 244 (AD-757-794), October, 1972.

The Enlisted Evaluation Center is the major operating element of the formal enlisted evaluation system. It was established in 1958 as a Class 2 activity of the Army. Its primary purpose then was to help the Army manage the proficiency pay program, which had been established in response to the recommendations of the Cordiner Committee--the Defense Advisory Committee on Professional and Technical Compensation--in 1956 and 1957.

Monetary incentives were one of the Committee's proposals designed to improve personnel retention and job motivation among trained technical specialties and, at the same time, stimulate higher quality performance among all enlisted personnel. Proficiency pay, as a concept, emanated from this Committee recommendation. However, an underlying principle of this concept was that proficiency pay must be directly related to the demonstrated level of proficiency and must be contingent upon periodic checks to ensure maintenance of that proficiency. The Army enlisted evaluation system was developed to meet this requirement.

ARMY ENLISTED EVALUATION SYSTEM

The enlisted evaluation system consists of two major components: (a) evaluation of the enlisted man's knowledge of the various duties that are required at his skill level in his MOS, as indicated by MOS evaluation tests and performance tests; and (b) evaluation of performance in the currently assigned duty position, as indicated by supervisory ratings on the enlisted evaluation report (see Figure 1). A rating system is applied to the scores obtained on these instruments, and it is used to compute a composite score for taking individual personnel actions. This MOS evaluation score indicates the individual's relative standing among those evaluated in the same MOS and skill level and in the same pay grade. It is used to verify MOS qualification, to assist in determining promotion eligibility, to award proficiency pay, to guide remedial training, and in a variety of other personnel actions.

PROFICIENCY TESTING "AREA SCORES"

The characteristics of the MOS proficiency testing program will not be detailed herein, but one particular aspect of the MOS proficiency test program directly relevant to this paper will be considered--the MOS "major area" scores. Each MOS evaluation test is organized into six to nine major areas; that is, six to nine subscores. The six major areas for an Infantry senior sergeant, for example, are weapons, tactics, field activities, unit defense, administration, and personnel accounting. Study references from Army regulations, pamphlets, field and technical manuals, and other manuals are coded to each of the major areas in an accompanying study guide so that each soldier can locate the printed materials upon which the test is based. He may study these reference materials to improve his knowledge and performance.

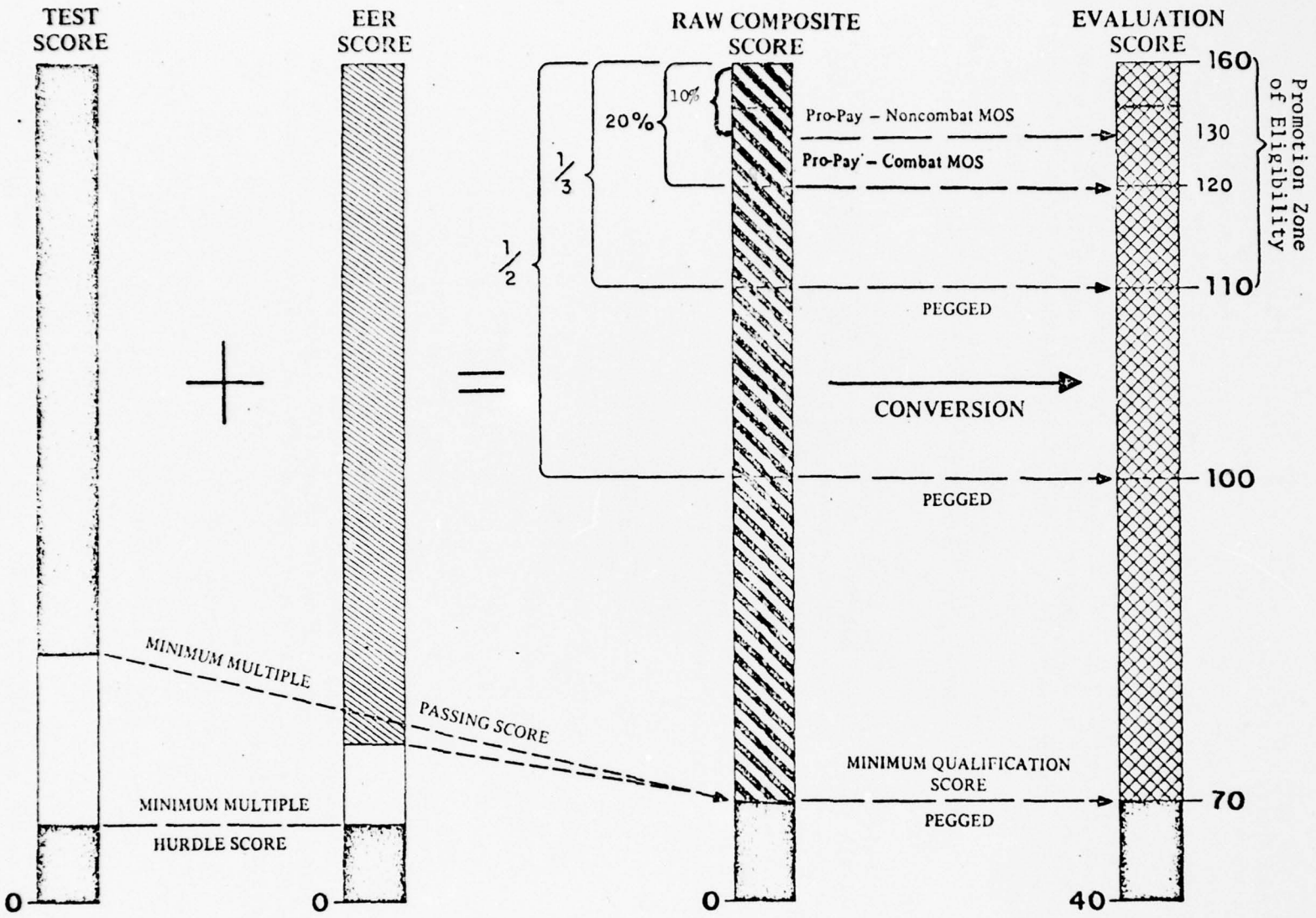


Figure 1. MOS Evaluation Score.[a]

[a] From U.S. Army Enlisted Evaluation Center, Briefing Supplement; Indiana, USAEEC, 1971.

The major areas are weighted according to the relative importance of the functions in the missions of all units that are authorized duty positions in the MOS skill level, and not on the basis of the time required to teach the subject matter in formal classroom courses nor on the basis of the number of personnel assigned or authorized for specific duty positions. The number of questions allocated to an area out of the total number of items in an MOS proficiency test indicates the weight assigned to that area. Subscores for these major areas are useful not only to the soldiers tested, who can use the information to improve their performance, but also to various headquarters, to centralized management programs, and to commanders for managing assignments and training programs.

DIFFERENT APPROACHES TO DUTY AREAS

Unfortunately, there is no consistent theoretical basis or consistent approach to the definition of these area scores by test developers, training personnel, or requirement planners (see Table 1). Some of the areas for which the Enlisted Evaluation Center has developed scores can be classified as duty areas. These area scores roughly correspond to subject matter areas within an MOS. They are used to provide information as regards an enlisted man's strengths and weaknesses in selected subject areas, and they are associated with specific subject matter references.

Table 1. Different Approaches to Duty Areas

Approach	Duty Areas
MOS proficiency tests	Selected by test developers
Army training schools	Identified by systems engineering of training
Requirement planners	Associated with additional skill identifiers

If subject matter references happen to be organized in terms of duty areas, it is easier to find the appropriate references that need to be studied; however, it is not essential. A sample list of major areas in an MOS is given:

1 Weapons
2 Tactics
3 Field Activities
4 Unit Defense
5 Administration
6 Personnel Accounting

The study guides list the regulations and technical manuals for the various areas, and no great amount of effort is needed to find the references

that correspond to the area in which a low score was received in the proficiency evaluation test (see Table 2).

Table 2. Sample Study Guide

References	Major Area
Army Regulations	
65-75	5
210-10	6
DA Pamphlets	
600-8	6
672-2	3
Field Manuals	
5-15	4
7-10	2

It should be clear that the rise of area scores for proficiency tests was an important development in the design of MOS proficiency tests. Such feedback systems are an integral part of any sophisticated testing program. What, then, would be questioned in the design of these area scores as they are used by the proficiency testing system? The point raised is not so much how the proficiency testing subsystem works, but the manner in which these area scores interface with other personnel subsystems in the Army.

The Army training schools, for example, identify major duty areas at great expense and with great difficulty, as part of their systems engineering of training process (see Figure 2). Systems engineering of training is a long, drawn-out procedure, involving detailed job analysis and the application of systems engineering principles and approaches in order to break a job into components and then select components for training. The job is also organized into "areas" when a Program of Instruction (POI) is prepared. There is no consistent relationship between these POI area scores designed by the training people and the area scores as used by the proficiency training people.

Requirement planners are also interested in a different kind of duty area. For example, the Army has additional skill identifiers (ASI) that are authorized for functional skills, which are not consistently required of all the job incumbents in an MOS. An example of such an additional skill identifier is the ability to maintain a specific type of system (e.g., maintenance on the Hawk Guided Missile Simulator, or the ability to work with specially trained scout dogs). These different approaches to duty areas are not necessarily incompatible with each other, but they are all different. When you have three different parts of the same organization -- and the Army is one organization -- using three completely different

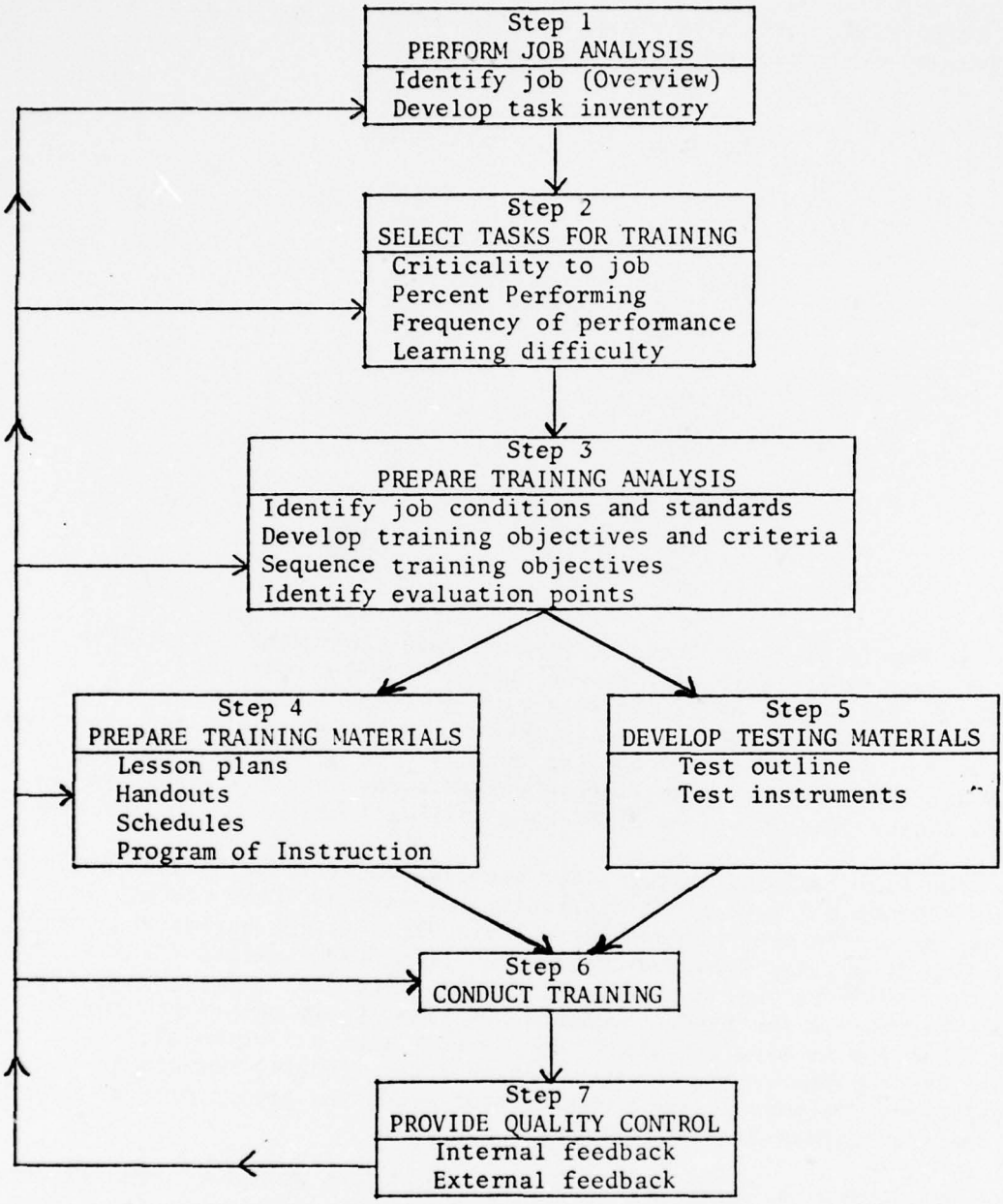


Figure 2. Simplified Flow Process of Systems Engineering of Training.[a]

[a]From Southeastern Signal School Briefing Supplement. Systems engineering of training at USASESS. Undated.

language systems to describe the same kind of work, it is likely that there will be some unnecessary duplication of effort. This could be avoided if a common language could be designed for all three parts of the organization (e.g., the Army) to use.

THE DUTY MODULE

A duty module is a group of occupationally interrelated tasks smaller than an occupational specialty. It is modular in the sense that it can be used as a plug-in unit to a variety of different occupational specialties. Table 3 defines the module group, number, and title. Table 4 shows an MOS duty module matrix for Army military occupational specialties. As one can see, a relatively small number of duty modules can account for seven different MOS. Notice that each of these military occupational specialties has demonstrable similarity with other MOS.

Table 3. Module Group, Number, and Title

A	ADMINISTRATION	
	A-1	Performs general administration at unit level
	A-2	Performs unit supervision and control of personnel
	A-3	Establishes and operates a unit mail room
	A-4	Types, files, and performs general clerical operations
B	TRAINING	
	B-1	Conducts or participates in unit and individual training
C	COMMUNICATIONS	
	C-1	Operates unit tactical communications equipment (excluding use of Morse code)
	C-2	Installs and maintains unit tactical wire communication systems
D	TRANSPORTATION	
	D-1	Operates unit combat support vehicles

(continued)

Table 3 (continued)

E TACTICAL OPERATIONS

- E-1 Prepares and employs maps, charts, and instruments in land navigation
- E-2 Engages enemy with tank and Armor vehicle mounted assault weapons
- E-3 Drives tanks and associated Armor combat vehicles
- E-4 Emplaces, reports, and neutralizes tactical obstacles
- E-5 Performs in mounted, dismounted, airborne or long-range patrols
- E-6 Engages enemy with mortars
- E-7 Participates in ground tactical operations as member of a maneuver unit
- E-9 Engages enemy in close combat with individual weapons and machine guns
- E-10 Engages enemy with recoilless rifles and direct fire missiles
- E-11 Functions under CBR warfare conditions

F STAFF MANAGEMENT

- F-1 Performs tactical operations support duties
- F-2 Performs tactical intelligence support duties

G MAINTENANCE

- G-1 Performs user maintenance on individual and unit equipment and weapons (excluding motor vehicles)
- G-2 Performs organizational maintenance on track and wheel vehicle mechanical systems
- G-3 Performs organizational maintenance on track and wheel vehicle electrical systems
- G-4 Performs maintenance administration

H FOOD SERVICE

- H-1 Establishes and operates a field mess
- H-2 Prepares and serves meals
- H-3 Operates a mess facility

I SUPPLY

- I-1 Establishes and operates a unit supply point

J PERSONNEL

- J-1 Initiates, posts, files, and retrieves information from personnel records
- J-2 Manages individual enlisted personnel and carries out manpower and personnel management programs
- J-3 Processes personal affairs actions for individuals

Table 4. MOS — Duty Module Matrix

Duty Modules[a]	Military Occupational Specialties						
	11B	11C	11D	11E	11F	11G	11H
A-1			x				
A-2	x	x	x	x		x	x
A-3	x		x	x			
B-1	x	x	x	x	x		x
C-1	x	x	x	x	x		x
D-1	x	x	x	x			
E-1	x	x	x	x	x	x	x
E-2				x			
E-3			x	x			
E-4	x		x	x			x
E-5	x	x	x	x		x	x
E-6		x					
E-7	x	x	x	x	x	x	x
E-9	x					x	
E-10							x
F-1			x	x	x		
F-2			x		x		

[a]See Table 3 for definitions of duty modules.

It is also possible to use duty modules to express personnel requirements. The list of work activities in Table 5 has not been formally approved, but it illustrates the type of approach that can be used. Given a data processing group, there is a need to supervise, to plan the analysis of the reporting, to keypunch, and so on. The number of full-time duty positions needed in the organization is ten. You can also specify the requirements in terms of the number of people qualified to perform each work activity. If you have a computer activity with a lot of night shift work, you are going to need at least three people who can supervise.

Table 5. Work Activity Requirements

Work Activity	Minimum Number of People Needed with This Skill	Number (or Proportion) of Full-Time Duty Positions
Supervision	5	1.00
Planning of analysis and reporting	3	.25
Receipt and verification of input data	3	1.00
COBOL programming	2	2.00
Keypunching	8	1.50
Computer operation, including peripherals	5	3.00
Interpretation of output	2	1.00
Preparation of reports	2	.25
Total Number of Work Activity Requirements	30	
Minimum Number of People Required		10.00

You may want one or two people for backup, in case of illness or vacations. Even though there may only be three duty positions involving supervision, you may want five people to be qualified as shift supervisors. Similarly, in planning the analysis and the reporting, you may want three qualified persons, but it is only a quarter-time job. In other words, defining work activities in terms of duty modules can provide a more efficient use of personnel with all functions covered, using a minimum number of personnel.

TESTS OF EFFECTIVENESS OF DUTY MODULES

Resource and requirement planning experts must first agree upon the qualification requirements. Secondly, compatibility with work practices in the field is involved. For this, actual survey data regarding the way in which tasks are assigned in the field can be employed. A third test is to evaluate the usefulness of the module in planning and evaluating the requirements for and performance of organizational units.

TEST 1

A first test is to ask experts to design job content modules. Typical job content modules are:

(1) Operates unit tactical communications equipment, excluding Morse code.

(2) Installs and maintains unit tactical wire communication system.

In this case, the people who are asked to operate unit tactical communications equipment are usually different from those who install and maintain it. Neither of these work activities, however, is a full-time position for anyone. These are modular things that can be assigned to different people.

TEST 2

A second test of duty modules is compatibility with assignment practices in the field. Some data have already been analyzed (see Figure 3). Data were used that were already in existence, and that had been collected with task inventories that were administered by the Army Office of Personnel Operations. The data base is called the Military Occupations Data Bank, or MODB. The original task statements in MODB are organized in terms of functional areas of performance (see the administration and training columns on the right side of Figure 3). The rows correspond to task clusters that were identified in an empirical clustering of tasks.

Number of Tasks in Cluster	Duty Modules: Admin. A 1	Admin. A 2	Admin. A 3	Admin. A 4	Trng. B 1	Trng. B 2	Co- 1	Admin. A	Training B	Co--- C
6	50	33						100		
9	44		33			22		77	22	
9	33			55				100		
4		50						100		
6						50			100	
13		45				15		85	15	
10					60	20		10	90	

Figure 3. Test 2, Compatibility with Assignment Practices in the Field (in Percent).

The Comprehensive Occupational Data Analysis Programs (CODAP) system, developed by the Air Force, was examined to determine it applicability.[4] This system was developed to cluster people rather than tasks, however. The Army Research Institute modified the CODAP system to the problem at hand.[5] People were not clustered together on the basis of their similarity in terms of task performance, instead, tasks were clustered together on the basis of the probability that the tasks would be assigned to the same people. On the left side of Figure 3, clustering is compared with some duty modules that were analyzed in terms of the "Obverse Cluster Analysis" system designed (see Table 3 for definitions of duty modules). The computer run suggested that there were seven task clusters in this particular group of tasks. The percentages shown in the duty module columns on the left side of Figure 3 indicate the percent of tasks in each of the empirically identified clusters that fall into duty module categories A-1 through B-2. The percentages shown on the right-hand side of the figure indicate the percent of tasks in the empirically identified clusters that fall into each of the administrative areas used to group tasks in the Military Occupations Data Bank.

It is important to note that the design of the duty modules is not complete, nor is the preparation of the task statements. These task statements will be revised to reflect additional job data and evidence concerning the organization and application of command authority in field units. Task statements will also be revised as the duty module system is further developed.

It does not necessarily follow that there will be complete agreement with the computer runs. For example, it is possible that the first three empirical task clusters would be considered as really one cluster rather than three. Before that conclusion can be made, many other occupational specialties, in addition to the one that these data are based upon, will have to be reviewed. Essentially, duty modules are derived from many different occupational specialties rather than just one, which is the case with this particular computer run. The decision as to whether these first three clusters are one cluster or two or three clusters must be based upon data in other occupational specialties as well.

TEST 3

A third test of the utility of duty modules is the applicability to the evaluation of unit performance in the field. Task checklist items have been

---

[4] Bottenberg, R. A., and Christal, R. E., "An iterative technique for clustering criteria which retains optimum predictive efficiency", The Journal of Experimental Education, 36, (4), Summer, 1968, 28-34.

[5] Edison, W., and Eastman, R. T., "Cluster analysis of occupational data with focus on task rather than people", U. S. Army Research Institute for the Behavioral and Social Sciences Research Memorandum, November, 1973.

devised based upon the duty modules. Each duty module consists of a collection of task statements. A study of Army Training Tests (ATTs) revealed that task statements could be converted into checklist items. This provides the means by which information can be collected about the performance of individuals during unit training tests. Separate scores are obtained for separate practice maneuvers. Separate scores can be obtained for movement, defense, and attack, as noted in Figure 4. Each exercise will have a scenario, the possibility of casualties, and so forth. Duty modules will be looked at during the appropriate phase of the Army Training Tests. They would not be tested during every single phase.

		Unit Test Phase		
Module	Task Checklist Item	Movement	Defense	Attack
C-1	Establish and operate field communications relay station		X	X
G-1	Make entries in equipment log books	X	X	

Figure 4. Step 3, Relationship to Unit Performance.

SUMMARY

These individual duty modules are designed to be derived from a variety of specialties rather than just one. They are economical in the sense that many different task inventories can be designed and jobs described with a small number of duty modules. They can utilize task inventory data already collected on assignment and work assignment practices in the field. They meet an apparent need for consistency at a level of generality between the MOS and the task, at an optimum level of detail. Ultimately, if duty modules can provide a language acceptable to both resources planners and requirements planners, they will facilitate communication and promote better matching of men and jobs.

PERFORMANCE STANDARDS AND SKILL LEVELS

It is not possible to talk about testing in terms of duty modules without first talking about testing standards. Performance standards in the Army are prepared systematically as part of the systems engineering of training process for the design of Army school courses. At one point in systems engineering, task and skill analysis sheets are prepared. These result in evaluation plans. It is a standard practice to indicate specific performance standards for training purposes for each terminal or facilitating objective on these evaluation plans (see Figure 5).

The evaluation plans only cover those areas of interest to the school, however. They usually do not cover all the tasks in an MOS, especially for the more advanced tasks, which the personnel are supposed to learn on the job. The Enlisted Evaluation Center, therefore, has to supplement this school-oriented information with other sources. Usually, they use the judgments of knowledgeable NCOs who have experience in the particular MOS, and who formulate proposals regarding what tasks are appropriately included in an MOS proficiency test.

1. Criteria for the Training Objective developed for the task:

   Action — Troubleshoot AN/TRC-24. (A-22-1)

   Condition — In addition to Standard Training Conditions, the student is given an AN/TRC-24 with one major component containing a DS part defect as well as OS-8, ME-30/U, TM 11-5820-287-12, TM 11-5820-287-34, and AN/TRC-24 Block Diagram.

   Standard — The student is qualified if, when given two defective AN/TRC-24s with a 2-hour time limit on each, he can isolate one of the defective parts.

Figure 5. Evaluation Planning Information Sheet.

The proficiency testing system in the Army, as it is presently organized, provides separate tests for each skill level within each MOS. Occasionally, one test may be used in two or three skill levels with different score requirements, but the principle is the same. One reason for providing separate tests for different skill levels is that an MOS is a broad collection of duty areas that cover many different duty positions. Providing separate tests for each skill level makes it possible to provide items that are more appropriate for the positions being filled by those who take the tests.

This skill level approach, which is not incompatible with the duty module concept that we have described, is appropriately used in connection with duty modules, and, further, it illustrates how useful duty modules can be to those who design tests. Table 6 lists the tasks for a duty module of patrolling, either mounted or dismounted. One's skill level is dependent upon whether one supervises a task, does the task and also supervises it, simply does the task, or whether one just assists in doing it. The three skill levels in the 11B MOS (Light Weapons Infantryman) are indicated. The skill levels are numbered 1, 2, and 4. There are only three skill levels in this particular MOS, so there is no skill level numbered 3. Table 6 indicates that people who are at skill level 4 are more likely to supervise. People who are at skill level 2 do not supervise, and are much more likely to assist somebody. People who are at skill level 1 carry out the orders and requirements of their superiors. This kind of information about skill level profiles could be extremely useful to Army organizations in designing proficiency tests.

Table 6. Duty Module E-5: Patrols, Either Mounted or Dismounted

Tasks	Performance Expectations for Skill Levels 1, 2 and 4 of 11B MOS[a]			
	Super-vise	Do and Supervise	Do	Assist
(1) Plan patrol operations			4	2
(2) Assemble, inspect, issue patrol order, and lead patrol			4	2
(3) Operate listening or observation post	4		1	
(4) Serve in combat patrols		4	2, 1	
(5) Serve in reconnaissance patrols		4	2, 1	
(6) Serve in ambush patrols		4	2, 1	
(7) Mark route or serve as guide for unit[b]			2, 1	
(8) Participate in air search operations or air delivered patrol		4	2, 1	
(9) Estimate charge, emplace and fire demolitions		4	2, 1	

[a]There are only three skill levels in the 11B MOS: 1, 2 and 4.

[b]Task No. 7 (when performed by an 11B MOS) is normally supervised by someone in another MOS.

ADVANTAGE OF A MODULAR APPROACH IN TEST DEVELOPMENT

Previously mentioned were the ways in which the Army military occupational specialties were modular. Refer to Table 4 and note that with this matrix one can account for many different tasks with a relatively small number of duty modules. Moreover, this matrix is just the corner of a much bigger matrix. To date, 31 enlisted duty modules have been developed; they completely account for 16 different 3-digit enlisted MOSs.

To appreciate the possible savings it is necessary to translate this information on duty modules and MOS into test items. Say that these 16 MOS would require 100 items apiece to account for them if they were developed independently by various test-developing agencies. That would make 1,600 test items, if you used an independent approach to test development. We estimate that it takes only 10 items apiece to describe a duty module. In other words, it is possible that 310 items can do essentially the same job as 1,600 test items. To be able to prepare 310 items rather than 1,600 items reflects a considerable savings. It is contingent upon defining modules that cut across, and have the same meaning in, different occupational specialties. In giving these figures, we have not discussed skill levels; but different tests for different skill levels would be required in both systems. Thus, multiply the number of items in our example by three or four to get the number of test items that would actually be needed by the people who design these tests and work with them.

MODULAR ARMY TRAINING TESTS (ATTS)

At the present time, the Army has several hundred Army Training Tests (ATTs) for use in evaluating the performance of organizational units. Each of these tests has a scenario and provisions for referees who are trained to follow people and take notes as regards their performance in the unit test. The question to be posed here is this: Would it be desirable for Army Training Tests for organizational units to be organized in the same modular fashion that has been proposed for individual proficiency tests?

PURPOSE OF ORGANIZATIONAL UNITS

Before further discussion of the feasibility of modular ATTs, the accuracy and consistency with which the intended purposes of organizational units have been specified must be considered. The Army has several different terms for describing the intended purposes of organizational units: a primary mission, some functions, and a capability.

A primary mission is defined as the principal purpose that an organization is designed to accomplish. The functions are the appropriate or assigned duties, responsibilities, missions, or tasks of an individual office or organization. A capability is the ability to execute a specified course of action. Further details will not be discussed except to state that after studying the various mission, functional, and capability statements in Army documents, it was concluded that the capability statement was the one that should be used as the basis for structuring organizational unit testing modules.

THE CAPABILITY STATEMENT

A possible capability statement for a theoretical unit is shown in Table 7. It is clearly possible to analyze a capability in terms of specific component functions and operational criteria, as shown in the table.

Table 7. Theoretical Capability of a Unit

1. Title:

   Transport supplies and resupply itself.

2. Essential Component Functions:

   a. Load, move, and unload unit loads of rations, POL, ammo, and repair parts.
   b. Repair minor vehicular failures enroute.
   c. If unable to make minor vehicular repairs, tow inoperable vehicles.
   d. Move unit loads on the road or cross-country.
   e. Pick up and issue supplies.

3. Minimum Operational Criteria:

   a. Sufficient vehicles on hand in condition to move unit load.
   b. Sufficient vehicles on hand in condition to pick up and deliver supplies.
   c. Trained drivers.
   d. Authorized maps and compasses on hand.
   e. Supply platoon trained as a team.
   f. Trained supply personnel.
   g. Trained vehicular and radio mechanics.
   h. Satisfactory status of equipment maintenance.
   i. Satisfactory completion of ATT and FTX.

4. Standards (To be developed):

   Includes minimum personnel, skills, operable equipment, and training necessary to be considered C-1, C-2, or C-3 as defined in AR 220-1. Standards below C-3 are C-4.

The capability of the unit is to transport supplies and to resupply itself. What is involved in this capability are the functions of loading, repairing vehicles, picking up initial supplies, and so forth. The minimum operational criteria are also indicated. A Department of the Army

study of output measurement conducted in 1968[6] shows the same conclusion that is arrived at here--that capabilities are a good way to structure and organize ATTs.

CRITERIA FOR MODULAR ATT EVALUATION DEVICES

A number of criteria for an improved system of evaluating performance in Army Training Tests have been formulated. Details are beyond the scope of this presentation, but the criteria and related recommendations are given.

- A series of clear, quantitative statements specifying the capabilities of a unit.
- A taxonomy of unit capability statements. If possible, these statements should be modular.
- Criteria of unit effectiveness based upon and relatable to the taxonomy of unit capabilities, and consistent with and relatable to criteria for the performance of individuals in the unit.
- Both kinds of criteria (individual and unit) based upon performance standards rather than relative standing in test performance.
- Varied performance standards depending upon situational conditions, such as terrain, percent casualties, and resource inputs.
- Aggregation of standards for organizational components and generation of an overall index for the unit as a whole.
- The emphasis in criteria statements upon end results rather than methods used to achieve the results.
- Output measures allowing for the possibility in evaluations of corrective measures that may have been taken by command personnel, and that would permit the unit to meet standards in spite of some departure from expected procedures.
- Scoring that provides specific information about the leadership of a unit.

---

[6] Department of the Army. Improvement of Output Measurement, Report of a Special Study by the Army Staff Coordinated by the Comptroller of the Army, January 1968.

- Scoring that provides comprehensive evaluations of the unit when tested as an entire unit.
- Quantitative weights assigned for satisfactory performance and deducted for inadequate performance that reflect the probable seriousness of the actions.
- A method for relating output measures to input measures logically.

SUMMARY

A series of clear, quantitative statements are needed to specify the capabilities of a unit. There is also a need for a taxonomy of unit capability statements. A taxonomy is a theoretically-based language that implicitly classifies or categorizes a capability statement at the same time that it describes it.

Capabilities should also be modular in nature; that is, the capability statement for an Infantry battalion should be the same, if possible, as a highly similar, closely related capability statement for an Armored Cavalry squadron. If you can design the capability statements thusly, and organize the testing accordingly, it is possible to design one modular unit training test component that would be useful for Infantry battalions and for Armored Cavalry squadrons. Examples of possible modules are night ground attack, retrograde movement, and stationary defense. Since many different capabilities of different kinds of battalions are common, having similar tests provide many economic features.

CURRENT WORK ON MODULAR ATTS

Having proposed the design of modular Army Training Tests and proceeding to design some, field survey work was conducted in conjunction with Army Training Tests at Fort Lewis, Washington, in August 1973, to pre-test the first versions of these modular evaluation devices. These devices and data have not been analyzed or evaluated as yet, thus comments are based upon what was learned during the design stages.

PERSONNEL CAPABILITIES

In designing these evaluation devices, two things quickly become apparent: (a) devices could not be designed for units based upon enlisted duty modules along; and (b) there was a need for officer duty modules. A contract already underway for field job analysis of about 200 officer jobs was used as the basis for designing initial duty modules for officer

jobs. Table 8 defines officer duty modules by area and module number. Table 9 accounts for all the major duties of an officer's job with just a few duty modules. Each of the positions shown is completely accounted for by the officer duty modules listed. (The letter refers to the "areas" from which the module was taken in Table 8, and the number identifies a specific module within that area.)

Table 8. Officer Duty Modules by Area

Area	Title	Number of Modules
A	Command Management, General Management and Administration	9
B	Personnel	4
C	Intelligence	5
D	Operations and Plans (Staff)	4
E	Organization, Training	3
F	Logistics (Staff and Consumer Units)	9
G	Communications and Electronics	2
H	Civil-Military Affairs	3
I	Comptrollership, Budget and Fiscal	2
J	Army Aviation	5
K	Research, Development, Test and Evaluation	2
L	Operations Research and Systems Analysis	1
M	ADP Management and Programming	1
N	Education, Instruction	2
O	Information Activities	1
U	Tactical Direction of Combat Units	5
W	Miscellaneous	9
X	Individual Functions and Special Qualifiers	4
FF	Logistical Services	9
HH	Supply and Maintenance Support Operations	9

Table 9. Application of Duty Modules to Officer Positions

Position	Duty Modules
Cdr., Infantry Rifle Co. CPT	A-1, A-3, A-4, A-6, A-8, E-1, F-1, X-1, X-2
Cdr., Reception Station Co. LT	A-1, A-3, A-5, F-1
Asst. Army Attache LTC	A-1, A-4, C-4

EQUIPMENT CAPABILITIES

In addition to information on the personnel capabilities of a unit, there was a need for detailed information about the capabilities of equipment. The need to think about equipment capabilities became apparent when preparing a unit capability table, which described different types of capability for each component in a platoon (Table 10).

Table 10. Equipment Capability Table for Armored Cavalry Platoon (TOE 17-107H)

Item of Equipment	Basis of Issue	Capability	Reference
1. Antenna (AT-784/PRC)	2-Scout Section 1-Rifle Squad	Determine the direction to a specific radio transmitting in the frequency range 30.0 to 75.95 MHz.	ST-24-18-1
2. Armored Reconnaissance Airborne Assault Vehicle (M 551)	3-Light Armor Section	Negotiate almost any terrain at speeds from 4 miles per hour in water to 43 mph on roads, including 7-foot spans, 33-inch vertical obstacles and 60% grades.	ST-17-1-1; ST-17-15-1; FM-17-36

It was not possible to describe the fire-power capabilities of a squad that was equipped with a certain type of machine gun, for example, without knowing what type of machine gun it was. One machine gun might be capable of a sustained firing of 40 rounds per minute, while another might fire 100 rounds per minute. The range of the first machine gun might be 6,000 meters, while the range of the second, faster machine gun might be 3,000 meters. Clearly, any quantitative statement of the capabilities of that squad, and hence the capabilities of the whole platoon, is greatly affected by which of the two machine guns is being used. Similar constraints upon unit capability statements are imposed by the kind of transportation that is available. This type of information has obvious implications as regards the capabilities of a unit, and must be incorporated or delineated in unit capability statements.

PRELIMINARY WORK ON EVALUATION DEVICES FOR ATTS

As stated earlier, one of the requirements for the individual enlisted modules is that they be meaningful in terms of unit evaluation procedures as well as individual evaluation procedures. However, caution must be exercised in the using or modifying duty modules at this stage. The task list for a given MOS may or may not adequately sample the tasks performed by incumbents. The task list also may or may not include all the tasks which make up the duty modules for that MOS. And finally, duty assignments may exist which are inappropriate to a duty position designation within an MOS.

APPROACH TO THE PROBLEM

The first approach was to develop examples of criterion behavior for a particular duty module, but it did not seem to be workable because too much territory was covered with one global rating.

A new approach was developed. Since each duty module is defined by approximately ten task statements, it was decided to design the evaluation devices in terms of checklist items based upon these task statements (Figure 6). The method of grouping the items was retained so that the devices were still organized in terms of duty modules. The overall performance of the unit could not be scored exclusively in terms of individual performance, however. Some unit performance indices were developed. An example of overall unit performance rating procedures is shown in Figure 7.

		ATT Phases					
Module Number	Checklist Item	Movement to and Occupation of Forward Assembly Area	Reconnaissance	Covering Force	Route and Area Security	Flank Guard and Economy of Force	Overall Rating
A-2	1. Replacements properly received and assigned 2. Losses and casualties properly processed 3. Leadership of platoon and squad NCOs 4. Duties of subordinates properly allocated						

Figure 6. An Approach to Unit Evaluation Devices Based upon Task Statements.

Activity Rated: Rifle Platoon	Score	Comments
Phase 1--Daylight Attack		
Proper actions and preparations in assembly area?		
Proper organization, formation, and dispersal?		
Platoon's use of cover and concealment?		
Firing on objective--good volume, well directed? (Scored only for live firing)		

Figure 7. Evaluation in Terms of Unit Performance As a Whole.

RECOMMENDED APPROACHES AND PLANS

APPROACH

Tests have usually been designed by taking a particular criterion situation and designing a test for that one situation irrespective of how it is structured for that particular case at that particular point in time. If, for example, a proficiency knowledge test is being designed for somebody who repairs automobiles, you determine exactly which tasks people in that MOS are supposed to perform, then you design the individual proficiency tests to measure knowledge of those specific tasks.

The same is true of unit training tests. You look at the capability statements, study the terrain in which the test is to be conducted, and then you design some tests for that particular type of unit in that particular situation. If you are dealing with an Infantry unit, you design an Infantry test; if you are dealing with an Armored Cavalry unit, you design an Armored Cavalry test. There is relatively little overlap or systematic utilization of materials prepared by other people for different kinds of units.

What is being proposed for individual testing is that personnel experts in the various services spend a lot more time dealing with a variety of occupational specialties rather than focusing upon one specialty at a time. If possible, they should change the way in which occupational specialties are defined, and define different occupational specialties in terms of a common set of duty modules.

What is being proposed for organizational units is that evaluation devices be designed so that different kinds of organizational units can be evaluated with the same kind of modular evaluation devices. The modular approach, in both cases, facilitates feedback regarding performance and makes the feedback to the individual or unit being tested much more meaningful. Both strong points and weak points are much better defined.

It is possible to argue that feedback about test performance should not be too specific. The Air Force, for example, does not provide information about area scores in its Skill and Knowledge Tests (SKTs), which are its equivalent of the Army's proficiency test. The reason given is that people would study for the test rather than study for the job (i.e., the Air Force would seem to be worried about the unintended consequences of providing information in terms of proficiency testing area scores). Available data suggest that area scores have been successful in the Army, however, and the unintended consequences simply have not occurred.

It would seem that, regardless of whether you provide feedback in terms of the modular scores that have been described, or whether you provide feedback in terms of areas defined by a committeed of NCOs or unit commanders, area scores are useful feedback and should be used.

The Army has already embarked upon a large-scale program to overhaul all of its unit training tests. The Army recently designed a systems engineering of unit training programs that is patterned after its systems engineering of training procedures for individual training. Several organizations are currently working on the redesign of unit training programs in terms of the Army's systems engineering of unit training programs. Thus, a number of new evaluation devices will definitely be designed. These new approaches to modular unit training tests have been discussed with those who are responsible for this type of work, and it is believed that they will consider a modular approach to ATTs as an alternative when the tests are revised.

USES OF DUTY MODULES

This paper has stressed the economy of duty modules in terms of test preparation costs and the importance of consistency in language. Other possible advantages are given.

(1) Duty modules can improve occupational research, its description, and utilization.

(2) They have the potential to reduce training time and lower training costs.

(3) They can provide a better use of individuals in assignment substitutions.

(4) They can simplify automated assignment and control procedures.

(5) They can improve proficiency evaluation.

(6) They can improve career guidance and planning.

(7) They can improve utilization of personnel at a local level.

(8) They can improve unit training evaluations.

PRESENT IMPLEMENTATION OF PROCEDURES

It would take many years before this kind of modular system could be established. If everyone agreed that it should be done tomorrow, it would still take several years before such a system could become operational. Nevertheless, the advantages of a modular approach to test construction are so great that people should start thinking about it now. The problem arises because test developers cannot proceed by themselves. The whole occupational structure needs to be revised so that occupational specialties would be defined in terms of duty modules.

Nevertheless, there are several procedures that could be started now. For example, change the way in which test design work groups are established. Instead of assigning test development to a group of experts in the same specialty, create a work group comprising representatives from several specialties, and ask that group to design a single test component that would be useful to them all. A similar approach (i.e., assigning groups of experts from different types of units) can be used with those who design unit evaluation devices. These preliminary steps, taken now, would have many immediate advantages, and would greatly facilitate a conversion to modular testing in the future.

THE GROWING DEMAND FOR HUMAN PERFORMANCE TESTING

-- J. E. GERBER, JR. --

ABSTRACT

This paper reports the human factors involved in moving United States Army Infantry School instructors away from norm-referenced, knowledge sampling testing and into criterion-referenced performance testing. It sketches the gradual convergence of and increased communication between subject matter experts (instructors) and test psychologists (quality controllers) from group orientation through stylized example to specific problems of performance testing, and compares the dynamics with those observed by educators in civilian schools.

INTRODUCTION

Parallel innovations have been occurring and recurring in different types of schools in separated parts of the country.

In April, 1973, Doll, Love and Levin (1) reported as follows on their experiences in implementing a new instructional model in Louisville elementary and high schools.

> "In Louisville, Kentucky, a brave new world of educational innovation was recently attempted. Bright ideas and good

intentions, a situation ripe for change, and a golly-gee, whiz-bang attitude at many levels of decision making augured well for its success. Education was to roll forward. History was to be made. But somewhere along the way the wheel had to be rediscovered and history not only was made, it was remade."

Two years ago, Kramer and Kneisel (2) showed you a model for course design as it was then being developed at the US Army Infantry School, and related to you a dynamic plan for imposing (2a) this model onto on-going courses of instruction. Their model is depicted on this slide.

(SLIDE 1 ON)

SLIDE 1. SYSTEMS ENGINEERING PROCESS

The two-headed arrow labeled quality control in the middle of the slide implies that quality control takes place at every step of the systems engineering process. Our concentration today will be on criterion testing.

(SLIDE OFF)

Thus, we at Infantry School had a model and an implementation plan for training adults to perform specific jobs.

Meanwhile, Getz and others (3), who were changing from a traditionally based to competency based programmed instructional model for teacher education at Illinois State University, provided the following admonition in early 1973:

> "While the rationale for any program can be developed logically, the actual implementation invariably creates many problems that cannot be anticipated... The results are rewarding, but adjustment by staff and students is slower and more agonizing than most would suspect."

PURPOSE

The purpose of this paper is to compare human experiences at the Infantry School with those at other institutions in changing instructional models; to focus on the impact of changing from norm-referenced tests of knowledge to criterion-referenced tests of abilities; to discuss some of the dynamics implied; and draw some inferences.

Implementation of the systems engineering model at the Infantry School has been a tremendous undertaking, both from the technical and the human side.

Systems engineering documentation for the first of some 22 active courses of instruction began in June 1968. We expect to have the final one completed in June 1974.

During this period through 1972, the emphasis was on documentation. Criterion testing was continued in those areas where it already existed, that is:

Pure performance testing, in which students performed job tasks against pass/fail criteria, continued in tests of physical fitness, weapons firing qualification, parachute jump qualification, and land navigation.

Performance testing under conditions simulating job conditions and against subjectively rated criteria was carried on in tactical map problems.

But, to a large degree, students were graduated from our courses on their ability to answer multiple choice items which required recognition, recall, or comprehension of facts or doctrinal principles, and which were graded "on the curve."

On 27 February 1973, we determined to make all tests performance oriented so that all would have the student apply job-required skills,

knowledges, and attitudes to an identified job task. While the reverberations of this decision were felt throughout the School, they were experienced most acutely by the instructor who relied most heavily on multiple choice items testing recognition or recall in his area of expertise. In order to relate subsequent events, observe the stylized functional diagram of the School organizational activities on this slide:

(SLIDE 2 ON)

SLIDE 2 STYLIZED ORGANIZATION & FUNCTION DIAGRAM

UNOFFICIAL

The Assistant Commandant runs the School.

The Director of Instruction is the staff officer in charge of designing and operating the resident courses. While he does not have direct, personal supervision over instructional department people, he does, nevertheless, direct or approve course actions, evaluate instruction and develop Programs of Instruction in the name of the Assistant Commandant. The systems engineers and course coordinators, working within the framework of a curriculum committee, are the principle architects of course design.

The test psychologists or quality controllers are the principal guardians of the instructional model presented earlier. Obviously, they, through their actions and judgments, are focal points for convergence of differences in rationale, procedure, and requirements.

The instructors, of whom there are approximately 1,050 in the School, have multifaceted roles under the systems engineering model. Lecturing a class of students, as he might have done daily in earlier times, is only one of the instructor's time consuming activities today. He works closely with systems engineers in documenting his area of expertise. He prepares his lesson plans with all that this entails. He works with quality controllers in his design and construction of tests and examinations. In fact, his role is not unlike that of the Illinois State University faculty member described by Getz and others (3a) and shown on the next slide.

(SLIDE 2 OFF)

(SLIDE 3 ON)

SLIDE 3 THE FOUR AREAS OF RESPONSIBILITY...

In addition to those marked on the slide, the USAIS instructor has other instructional duties which are physically and mentally demanding such

as field problems and weapons firing. The instructor faces massive problems of priority.

(SLIDE 3 OFF)

The multiple demands upon the instructor's time and attention account in part for the spectrum of response to the requirement for performance testing. I emphasize the word spectrum, for a spectral response did occur, varying from, "Great, I'm doing real world performance testing already," to "No way." We applaud the former and the many who flocked to it in short order, but it was those on the "No way" end of the scale who required the most attention.

This paper deals with dynamics of change. If a given reaction or attitude was encountered at all, it is discussed here for its dynamics, not for its frequency.

METHODOLOGY

Quality controllers conducted a series of free-wheeling meetings with groups of 10 to 20 instructors at a time to surface problems and to establish the kind of dialogue and openness which Doll and his Louisville associates (1a) believe essential. The problems surfaced bore a marked resemblance in kind if not in content to those reported by Gross

and his Harvard colleagues (4) in their attempt to implement a modular change in a small elementary school. These were:

1) Staff resistance;

2) Misconceptions of the model and roles within it;

3) Lack of expertise;

4) Lack of materials and resources;

5) Incompatibility with the model.

It became apparent to us that not only were there some problems of implementation, many of them were problems that had been raised elsewhere some years before as will be related in part below.

We determined that, having uncovered our own choke points or bottlenecks, we should press on while heeding the following speculation of Doll and his associates (1b):

> "...the wheel <u>does</u> have to be eternally re-invented...; unless you go through every agonizing step yourself, whatever you accomplish ultimately will be superficial and unimportant... Each of us must discover for ourselves where it hurts the most, but the experience of others may be helpful in learning how to cure the wound."

In this endeavor, we determined to work toward changing old concepts and practices into new ones using a proactive-retroactive transitional rationale described by the present author (5) in 1964. On the way we noted that Popham (6), at least as early as 1968, warned the educational community that many arguments against "stating instructional goals in terms of measurable learner behaviors" were being voiced. Popham felt that all these arguments were invalid, so, in order to refute them, he codified these arguments into eleven reasons why a behavioral model was or is said to be improper. Innovators at the Infantry School, at Illinois State University, (3) and Louisville Public Schools (1) encountered several of these "reasons" when changing from one instructional or testing model to another. Therefore, it seems appropriate to compare experiences and management techniques at these three schools, using the negative arguments as focal points. Perhaps others planning to roll toward pedagogical innovations may profit by our re-invention of the wheel.

Popham's negative reason number 1 is shown on this slide.

(SLIDE 4 ON)

SLIDE 4 NEGATIVE REASON 1 TRIVIAL LEARNER BEHAVIORS

(SLIDE OFF)

A comparison of experiences and management at the three schools which I have related to Reason 1 is shown on this slide:

(SLIDE 5 ON)

SLIDE 5 REASON 1: "TRIVIA"

We see that USAIS and ISU (3b) encountered the problem, whereas Louisville (1c) apparently had a more positive experience even though some of the sample objectives presented seem to be taxonomically low.

In our experience, operationalizing at the lowest level was easy. It took the form of, "At the conclusion of this instruction, you will be able to list, state, compare, etc." The student needed only sufficient mental capacity to recognize, recall or understand instruction in order to pass tests.

(SLIDE OFF)

The mental capacities and the elicited behaviors are at the lower end of Bloom's (7) taxonomy of educational objectives illustrated as a stairstep or hierarchy by this slide:

(SLIDE 6 ON)

SLIDE 6 HIERARCHY OF MENTAL REQUIREMENTS...

This slide, which is a modification of Towne's (8) illustration of Bloom's taxonomy, compares mental skills with job performance. Our difficulty arose in trying to raise the taxonomic level from the first two levels--knowledge and comprehension--to the applicatory level and higher. In the case of knowledge, we want to say, "Given a situation such as you are likely to encounter in your future job, you will be able to apply your newly acquired knowledge well enough to analyze your situation, synthesize all pertinent factors, make an evaluative judgment, and behave appropriately."

We found solidly imbedded defense of subject matters as having intrinsic value--knowledge for knowledge sake--regardless of the job for which the student was being trained. The position was commonly stated as, "The student just needs to know this."

(SLIDE OFF)

In managing or changing this position, we had some success in face-to-face encounters by asking, "Why?" Success came most often when the instructor could reply, "Because, if the student is ever in BLANK situation, he would have to do thus and so."

In this way the barrier to higher order operationalizing was broken for that particular instructor for that particular bit of information. If the

answer had been generated more affectively than cognitively, the technique succeeded where logical argument and cogent reasoning had failed. Some instructors were more concerned with the minutia of goal achievement than with the achievement itself. It was more important, for example, for the student to know the fact that a hole must be dug to a certain depth than it was to specify acceptable behavior for determining hole depth. This attitude was amenable to suggestion that a slight change of teaching strategy would present both the doctrine and the terminal behavior.

Popham's reason 2 (6b) is shown on this slide:

(SLIDE 7 ON)

SLIDE 7 REASON 2: PRESPECIFICATION OF EXPLICIT GOALS...

(SLIDE OFF)

The comparison of the three schools is shown on this slide.

(SLIDE 8 ON)

SLIDE 8 REASON 2: "UNEXPECTED OPPORTUNITIES"

We see that experiences among the three schools differ. The differences may stem from the different frames of references, i.e., adult,

face-to-face instruction at USAIS; adult self-instruction at ISU (3a); and elementary child face-to-face instruction at Louisville (1d). In our experience, serendipity in the classroom is rather tightly controlled both administratively and by the students themselves. There is a growing demand by students to be informed of the precise goals of impending instruction, aided by administrative demand for precise statement of both immediate and ultimate goals. Novel, spontaneous student responses in class are encouraged, but the class members themselves chasten the serendipitous student. Thus, even though we conduct experimental instruction, by the very nature of our training task we suppress serendipity by choice and relieve boredom by a multi-media approach. Tests, obviously, must test the training objectives, not the "old war story" which might be used to illustrate solution to a job task dilemma.

(SLIDE OFF)

Popham's reason 3 (6c) is shown on this slide.

(SLIDE 9 ON)

SLIDE 9 REASON 3: BESIDES PUPIL BEHAVIOR CHANGES...

For the present purposes, I shall compare and discuss only changes in the professional staff.

(SLIDE OFF)

Comparisons of staff attitudes and behaviors are shown on this slide.

(SLIDE 10 ON)

SLIDE 10 REASON 3: "ATTITUDE CHANGE"

The reaction to proposed change was spectral at all three schools, so we may concentrate discussion on the resistance or opposition end of the scale.

Some members of the USAIS professional staff expressed themselves as follows: "We have gotten along without job knowledge performance testing for nearly 200 years, and we have won many wars during that time; why, all of a sudden, do we have to change now?"

Those instructional staff members who held this or related views presented the greatest challenge because, unlike the Louisville plan (1e), our implementation plan provided no immediate escape for those opposed to change. Moreover, neither we nor the Louisville schools (1a) had the option of closing down while we sorted ourselves out. Some viewed systems engineering and resultant requirements for performance testing as change for sake of change, others as schemes

dreamed up independently by quality controllers who, as a result, sustained a certain amount of personal denigration to which Doll (1f) also refers. Some viewed the changeover as direct attacks on their areas of expertise, pedagogical approaches, or organizational division of labor. Some saw the change as a plan worthy of resistance or subversion, as the Louisville administrators foresaw (1g).

Even so, these members of the staff responded positively to three kinds of suggestive challenge in the military setting:

First--Look at the kind of testing you are doing now. If your life depended upon the ability of one of your students to perform in your area of expertise, would you say he was qualified on the basis of the test you give now?

Second--If you cannot show how your subject applies to the graduate's job, why not delete your subject from that course?

Third--If application of the facts you teach is taught by someone else, why not stop trying to test application now?--Why not get together with the other instruction and test later?

(SLIDE OFF)

Popham's reason 4 (6c) is shown on this slide.

(SLIDE 11 ON)

SLIDE 11 REASON 4: "MEASURABILITY IMPLIES BEHAVIOR..."

(SLIDE OFF)

Comparison of the three schools is shown on this slide.

(SLIDE 12 ON)

SLIDE 12 REASON 4: "DEHUMANIZING"

The ISU experience (3b) makes the case for criterion testing of overt behavior at the teacher college level.

The Louisville experience (1h) makes a case for contingency planning under a whole person interaction concept.

As previous remarks on Infantry School experience with Reasons 1 through 3 imply, we did not run head-on into cries of reductionism, robotism, or similar charges of dehumanizing either instructors or students. In fact, part of our underlying appeal for performance oriented testing has been on grounds of human interdependence. We asked for and received imaginative replies to this generalized question: "What is the minimum behavior you will accept now from the

student that will convince you that he can perform in your area of expertise?" The replies were translated into criterion statements of desired behavior and hence into performance tests.

In the so called "soft skills" areas--the leadership and attitude areas addressed by the present author in 1972 (9)--subject area specialists working to the point of exhaustion with test psychologists progressed from a condition of no validating tests, through multiple choice knowledge regurgitation tests, to case study type situational judgment tests of application of principles and knowledge. In addition, they have developed leadership peer rating scales on the premise that leadership qualities depend, at least in some degree, upon the perceptions of those who are to be led. In addition, they have developed a judgmental rating chart for race relations instructors to use in judging fitness of students to graduate as group facilitators. Significantly, here, according to Rogers (10), one of the appropriate behaviors is keeping quiet, remaining still, and allowing contentious group members to relate to the group.

Experiences at the three schools seem to point to the desirability of stating performance objectives in advance and testing them by some

means other than the objective type, norm-referenced paper and pencil test.

(SLIDE OFF)

Popham's reason 5 (6d) is shown on this slide.

(SLIDE 13 ON)

SLIDE 13 REASON 5: "IT IS SOMEHOW UNDEMOCRATIC..."

(SLIDE OFF)

Comparison of the three schools is on this slide.

(SLIDE 14 ON)

SLIDE 14 . REASON 5: "UNDEMOCRATIC"

This charge does not appear to be a problem at any of the three schools (1g).

Our students are encouraged to demand to know in advance the objective of an impending lesson or exercise. Most of the students are in courses of instruction by choice, planning their futures around their to-be-acquired ability to perform in a given military job field upon graduation. They tell us, sometimes vociferously, when an examination fails to meet their expectations to allow them to demonstrate their

ability to perform future job tasks. True, there are some who complain about having to make decisions on examinations, but not many.

(SLIDE OFF)

Popham's reason 6 (6d) is shown on this slide.

(SLIDE 15 ON)

SLIDE 15 REASON 6: "THAT ISN'T REALLY THE WAY TEACHING IS..."

(SLIDE OFF)

The comparison is on this slide.

(SLIDE 16 ON)

SLIDE 16 REASON 6: "REALISM"

Discussion of our experience relates to that under Reason 1, above. Our staff and faculty was educated, in the main, under a liberal arts system which emphasized:

1. Acquisition of knowledge, to be applied--perhaps--in that great world beyond;

2. Independent research, in which knowledge, vicariously gained, was self satisfying and possibly stimulating to yet additional research.

Relatively few, including those of the United States Military Academy, have been trained and educated to perform a specific job. Now, suddenly, these graduates, who are now faculty members, are required to write and teach toward behavioral objectives and to devise tests requiring appropriate student behavior at a given criterion level under controlled conditions.

They were not taught like that.

Teaching was not like that.

Teaching is now like that.

And, as we are seeing, the change from one frame of reference to the other spawns collisions.

We have managed the overall change through successive approximation. Quality controllers at first urged, helped, required instructors to use action verbs in their teaching or training objectives, and later required more and more specificity of the conditions and standards of performance to be elicited from the student.

Quality controllers are continuing to help instructors develop tests of identified job tasks which are as closely related as possible to job actions, conditions and standards. While some of the tests

remain in the multiple choice mode, in which the student selects one of the stated actions as his solution, we are pulling away from these as we develop the capability, resources, strategies, and instruments to observe and grade outdoor and indoor situational performance.

(SLIDE OFF)

Popham's reason 7 (6e) is on this slide.

(SLIDE 17 ON)

SLIDE 17 REASON 7: "IN CERTAIN SUBJECT AREAS..."

(SLIDE OFF)

The comparison is shown on this slide.

(SLIDE 18 ON)

SLIDE 18 REASON 7: "HUMANITIES"

Obviously, all three schools encountered the problem (1d, 3b).

In our early experience, an instructor would feel that systems engineering and performance testing were fine for everyone else except him, because of the nature of his subject area. A few went to great lengths to "prove" that performance testing their areas was impossible. Their cases were strengthened, spuriously, by early inability of test

psychologists who were unfamiliar with specific doctrine and subject areas to come up with instant test items, schemes or scenarios.

We managed this by continuing to iterate performance testing as the goal under a stance we termed "urgent evolution, not instant revolution"; by frequent, open, and sometimes protracted test item development sessions between instructors and test specialists. These took on a Socratic aura in that the testers continually asked questions of a "could you do this" type until the instructor himself eventually constructed the test. We, in the words of Doll and his associates (li):

> "held firmly to the notions that people support what they help create, that people affected by change must be allowed active participation and sense of ownership in the planning and the conduct of the change."

(SLIDE OFF)

We believe that the approach was successful because both the test psychologist and the instructor gave freely of themselves in order that the instructor might succeed. If this sounds humanistic, it is so intended, for it fits rather well with Buehler and Allen's (l l) discussion of Bugental's humanistic ethic:

"Integral to his scheme is his emphasis that the humanistic.. educator—person hopes that via his interventions and interactions he himself and the individual with whom he is interacting will emerge from that experience as societal change agents themselves."

This Socratic tedium is still going on, but, as Bugental would have it, the conferences are more and more frequently now between instructor and instructor in the questioner and responder roles.

At this point I should say that some of our best constructed paper-and-pencil tests are the military tactics planning and movement exercises laid on terrain maps. However, these tests have no precise solutions. They are graded subjectively and serially by a team of 3 tacticians who use their own judgments as to how good a tactical move the student made and how good the student's justification of the move is. Failure rates are traditionally high even though norm-referenced grading has not been completely eliminated. Quantification of tactical judgments in order to determine why this is so may indeed prove a difficult task.

Popham's reason 8 (6e) is shown on this slide.

(SLIDE 19 ON)

SLIDE 19 REASON 8: "WHILE LOOSE GENERAL STATEMENTS OF OBJECTIVES..."

(SLIDE OFF)

The comparison is on this slide.

(SLIDE 20 ON)

SLIDE 20 REASON 8: "INNOCUOUS OBJECTIVES"

As skills in identifying and writing behavioral objectives are being developed, it is simply axiomatic that poor, weak, or innocuous objectives will be included (1d, 3b). The prospect is allowed for in the all pervasive quality control provision of the systems model used at all three schools (1, 3). Infantry School instructors, and probably others as well, found that trying to construct performance tests on innocuous or poorly constructed behavioral objectives was essentially impossible. This impasse also held true when constructing tests on objectives which did not relate to the job for which the student was being trained.

During conferences with an instructor who was having great difficulty applying his subject area content to the course graduate's job, the test psychologist would help the instructor define the problem but would never point the finger or try to dictate course or lesson content. As a result, instructors sometimes drastically revised objectives and instruction in order to teach and test critical job skills.

(SLIDE OFF)

Popham's reason 9 (6f) is on this slide.

(SLIDE 21 ON)

SLIDE 21 REASON 9: "MEASURABILITY IMPLIES ACCOUNTABILITY ..."

(SLIDE OFF)

The comparison is on this next slide.

(SLIDE 22 ON)

SLIDE 22 REASON 9: "ACCOUNTABILITY"

The Louisville (lg) elementary school innovation has an accountability scheme somewhat like that of our own adult training school wherein the learners, the faculty members, the administrators, and the parents or, in our case, employment market controllers have a voice in kind and standards of learner performance.

Illinois State (3a) reported a fragmented accountability framework which is also similar to ours as illustrated on previous slides (Slides 2, 3).

Our faculty members do not fear accountability. In fact, most demand it. Some argued tenaciously for the right to give tests of knowledge and comprehension, saying, "I want to know that the student knows what I taught him. If he fails the course later, it will not be because I didn't teach him my subject."

This stance was understandable. We managed it by saying in effect, "You may use whatever instructional strategy you desire, including tests of factual knowledge. But when the student is certified in your area of expertise, he must be certified on a performance test. If you teach only facts, and someone else teaches application of the facts, the test must be on the application. Results of your in-class information or practice tests can be given to the student's faculty advisor for consideration if the student's progress appears in doubt."

We found one area which might have been fear of accountability but was more likely a liberal arts compassion for the students. The instructor proposed an examination of several problems wherein the student could

omit one or more. This plan was abandoned when test specialists pointed out that it did not assure testing of all the training objectives in the instructor's area of expertise.

(SLIDE OFF)

Popham's reason 10 (6g) is on this slide.

(SLIDE 23 ON)

SLIDE 23 REASON 10: "IT IS FAR MORE DIFFICULT..."

(SLIDE OFF)

The comparison is on this slide.

(SLIDE 24 ON)

SLIDE 24 REASON 10: "GENERATE OBJECTIVES"

Whereas the ISU self-instruction packages (3b) are said to contain all the competencies considered necessary for teacher training, we noted previously that not all the higher order competencies were specified.

At Louisville (1h), workshops in systems language and in writing behavioral objectives were run initially but were not continued after the first year. Instead, the school entered a plateau phase--a time during which concepts, ideas, and motivations hopefully are being internalized.

In our experience, some instructors raised the question as to who properly should write the training objectives. Outside agencies? Systems engineers? Instructors? Some instructors believed very strongly that their job was to teach, and nothing else. Some felt that, because of their junior officer grades, they were singularly unqualified to construct and conduct tests.

We managed to gain their cooperation in helping to write objectives by stressing:

1. Need to certify graduates;

2. Graduate capability to perform in the instructor's area of expertise on the job;

3. The logically inseparable relation of teaching and testing;

4. Teaching and verifying or testing as two aspects of the same behavioral modification process;

5. Qualification to teach carries with it the obligation to test;

6. Writing the action, conditions and standard of the training objective as tantamount to writing the test.

Here again, the approach was by successive approximations to the ideal product.

(SLIDE OFF)

Popham's reason 11 (6g) is on this slide.

(SLIDE 25 ON)

SLIDE 25 REASON 11: "IN EVALUATING THE WORTH OF INSTRUCTIONAL SCHEMES..."

(SLIDE OFF)

The comparison is on this slide.

(SLIDE 26 ON)

SLIDE 26 REASON 11: "UNFORESEEN EVENTS"

Unforeseen events touched all three schools. Mentioned earlier were Louisville's (1h) underemphasis on objectives for classroom management and Illinois State's (3c) faculty and students' slow, agonizing adjustment to competency-based teacher education.

In our experience, the role of the little 10 minute quiz became the focus for philosophical clash. On the one hand were those who wished to verify learning of facts without application of those facts to the learner's future job. There were also those who viewed education

as punishment and the quizzes as a whip, an outside motivator or disciplinary measure necessary to assure that students read their homework assignments. These proponents would penalize students, academically, for poor answers to pop quiz questions.

The problem was overcome by extensive revision of school testing policy and by separating these brief tests into two categories:

1. Instructional techniques, progress checks or feedback devices;

2. Verifying techniques, or performance tests of ability to perform job tasks.

(SLIDE OFF)

SUMMARY

Hopefully this paper has answered, at least in part, the following questions about criterion-referenced measurement posed by Dziuban and Vickery (12) in February 1973:

"How are teachers to make the transition from the more traditional practices and what are the consequences?"

"Can present instructional material be adapted to criterion-referenced measurement?"

"Will a new system ultimately result in substantial additional demands upon teachers, many of whom are presently operating on overcrowded schedules?"

CONCLUSIONS

I believe that we can draw the following general conclusion from the experience at the various schools as presented and discussed here: Experiences with radical change from one frame of reference to another have been about the same at USAIS as they have been at other schools that we know about. In the words of Doll and associates at Louisville (1b), this conclusion is stated as follows:

"...problems which have arisen in bringing about educational reforms..exemplify and reinforce findings of previous research..."

CLOSE

This paper is not intended to malign anybody. Instead, and again I use the words of Doll and his associates (1j):

"It ... is intended as a salute to people who tried boldly to make a significant impact on problems which bedevil every large ... school system. Their mistakes stand out clearly

because they attempted more than anyone else to date. Their successes may seem unduly minor because they attempted to conquer a whole mountain range and now only control some of the peaks. But they have at least gotten past the foothills, and from the peaks they can see more clearly than before. They can continue to climb. Hopefully, they can help those of us at base camp who may try to join them."

REFERENCE

1. Doll, Russell C.; Love, Barbara J.; and Levine, Daniel U. "Systems Renewal in a Big-City School District: The Lessons of Louisville." Phi Delta Kappan, April, 1973, pp 524-534.

1a. Ibid, p 527.

1b. Ibid, p 530.

1c. Ibid, pp 526, 7.

1d. Ibid, pp 527, 8.

1e. Ibid, pp 525, 7.

1f. Ibid, p 528.

1g. Ibid, p 525.

1h. Ibid, pp 526-8.

1i. Ibid, p 529.

1j. Ibid, p 524.

2. Kramer, Bryce R., and Kneisel, Richard S. "Systems Approach to Evaluation and Quality Control of Training." US Army Infantry School, Fort Benning, GA 31905. Presented to Military Testing Association Conference, Washington, D. C., September, 1971.

2a. Ibid, p 6.

3. Getz, Howard; Kennedy, Larry; Pierce, Walter; Edwards, Cliff; and Chesebro, Pat. "From Traditional to Competency-Based Teacher Education." Phi Delta Kappan, January, 1973, pp 300-302.

3a. Ibid, p 302.

3b. Ibid, p 301.

3c. Ibid, pp 301, 2.

4. Gross, Neal; Giacquinta, Joseph B.; and Bernstein, Marilyn. "An Attempt to Implement a Major Educational Innovation: A Sociological Inquiry." Cambridge. Harvard University. Center for Research & Development on Educational Differences, 1968, Chap 6.

5. Gerber, J. E. Jr. "Proactive and Retroactive Effects in Programmed Learning." In Ofiesh, G. D., and Meierhenry, W. C. (ED), Trends in Programmed Instruction. Washington, NEA, 1964, pp 232-234.

6. Popham, W. James. "Probing the Validity of Arguments Against Behavioral Goals." In Kibler, R. J., Barker, L. L., and Miles, D. T., Behavioral Objectives and Instruction. Boston, Allyn & Bacon, 1970, pp 115-124.

6a. Ibid, p 116.

6b. Ibid, p 117.

6c. Ibid, p 118.

6d. Ibid, p 119.

6e. Ibid, p 120.

6f. Ibid, p 121.

6g. Ibid, p 122.

7. Bloom, Benjamin S. Taxonomy of Educational Objectives, Handbook I; Cognitive Domain. New York, McKay, 1956.

8. Towne, William B. Sr. "Differentiating Between Situation-Oriented Items and Setting-Oriented Items in MOS Proficiency Testing." 5 Seymour Pl., North Augusta, S. C. 29841 (Unpublished manuscript, personal communication), October, 1972, p 3.

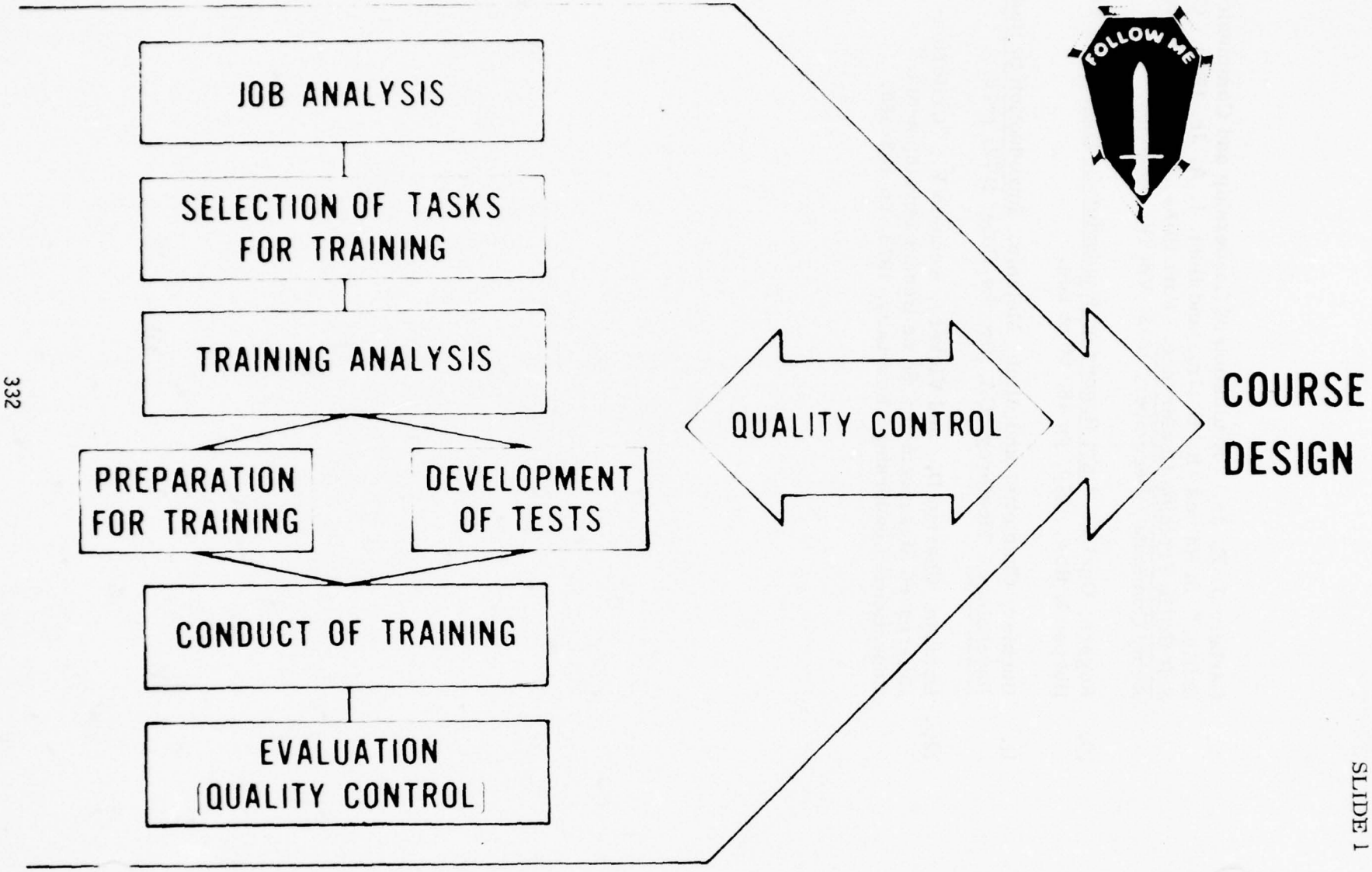
9. Gerber, J. E. Jr. "Evaluation of Leadership and Communicative Skills." In Haines, R. E. Jr., and Hunt, I. A. Jr. (ED). CONARC Soft Skills Training Conference. Fort Bliss, TX, US Continental Army Command, December, 1972, Vol IV, pp 28-34.

10. Rogers, Carl R. Carl Rogers on Encounter Groups. New York, Harper & Row, 1970, pp 48, 66 et seq.

11. Buehler, Charlotte, and Allen, Melanie. Introduction to Humanistic Psychology. Monterey, CA, Brooks/Cole, 1972, p 78.

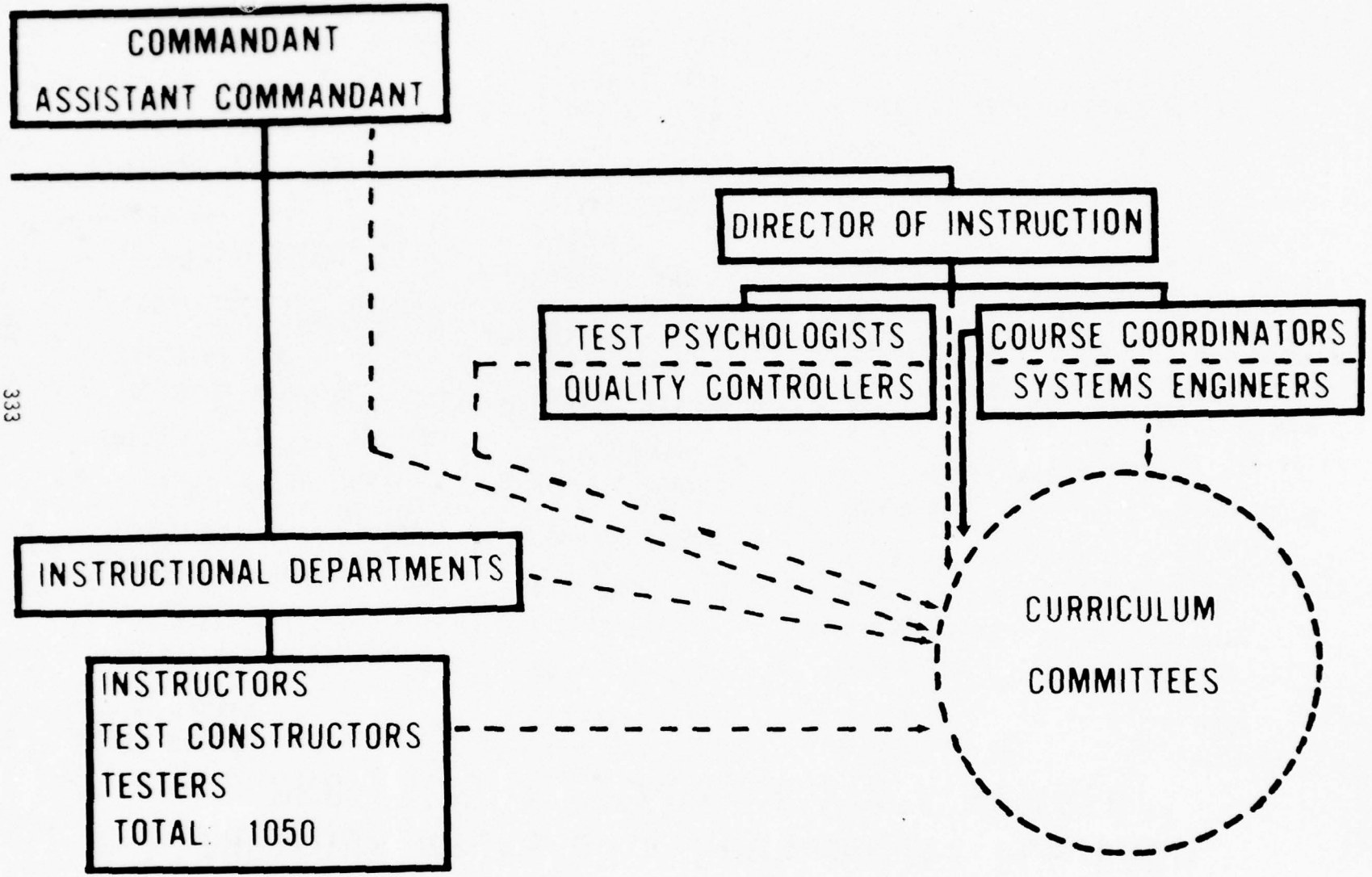
12. Dziuban, Charles D., and Vickery, Kenneth V. "Criterion-Referenced Measurement: Some Recent Developments." Educational Leadership, February, 1973, pp 483-486.

SYSTEMS ENGINEERING PROCESS



SLIDE 1

STYLIZED ORGANIZATION & FUNCTION DIAGRAM



BEST AVAILABLE COPY

THE FOUR AREAS OF RESPONSIBILITY FOR EACH PROFESSIONAL SEQUENCE STAFF MEMBER

A JOB WITHIN THE SYSTEM	B INDIVIDUAL TEAM PARTICIPATION TASKS	C SEQUENCE GUIDE REVISION	D COMMITTEE MEMBER
* 1 MAINTAIN LIBRARY MATERIALS	* 1 ATTEND TEAM MEETINGS	* 1 BUILD NEW PACKAGES OR REVISE OLD PACKAGES	* 1 ATTEND SUBJECT MEETINGS
. COORDINATE JUNIOR PARTICIPATION EXPERIENCE	* 2 CORRECT PAPERS	* 2 BUILD NEW TESTS OR EXPAND OLD TESTS	* . PROVIDE FEEDBACK FOR CENTRAL POLICY COMMITTEE
* 3 PREPARE INSTRUCTIONAL SESSION SCHEDULES	* 3 OFFER FEEDBACK TO STUDENTS	* 3 BUILD NEW MEDIA OR REVISE OLD MEDIA	* 2 VOTE ON ACCEPT ANCE OF NEW PACKAGES AND TESTS IN AREAS OF EXPERTISE
* 4 SUPERVISE MICRO-TEACHING LAB	* 4 PROVIDE PERSONAL AND ACADEMIC ADVICE	4 ETC	4 ETC
5 UPDATE MESSAGE SERVICE	* 5 INITIATE AND CONDUCT INSTRUCTIONAL SESSIONS		
6 EDIT PROFESSIONAL SEQUENCE GUIDE	* 6 SERVE AS TEAM CHAIRMAN ON A ROTATING BASIS		
7 OVERSEE THE SURVEIL LANCE SYSTEM	ETC		
* 8 OPERATE THE TESTING PROGRAM			
* 9 COORDINATE CREDIT NO CREDIT FOR GRADE CARDS			
10 ETC			

* APPLIES ALSO TO USAIS INSTRUCTOR

NEGATIVE REASON 1.

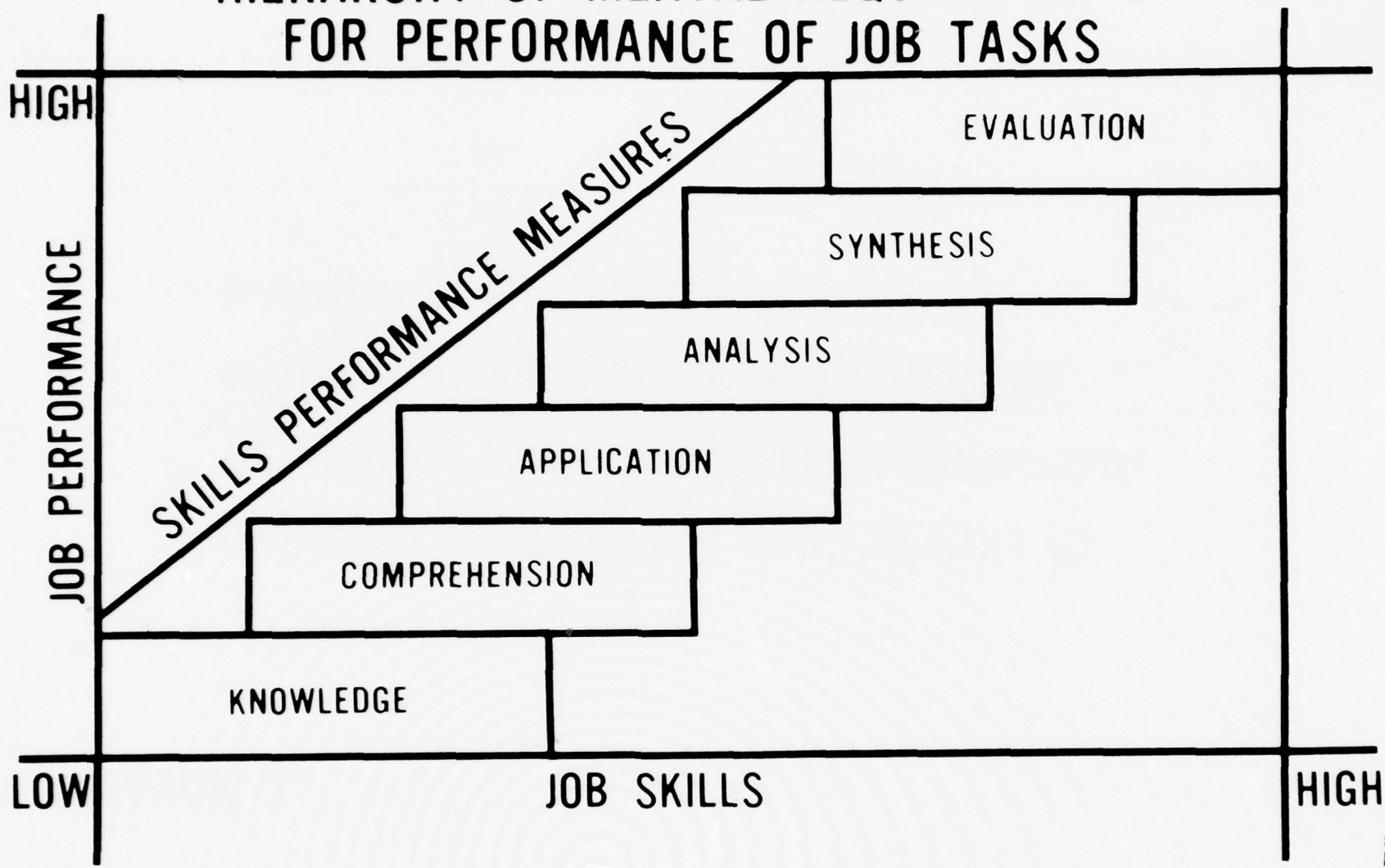
"TRIVIAL LEARNER BEHAVIORS ARE THE EASIEST TO OPERATIONALIZE, HENCE THE REALLY IMPORTANT OUTCOMES OF EDUCATION WILL BE UNDEREMPHASIZED."

BEST AVAILABLE COPY

REASON 1 "TRIVIA"

EXPERIENCE AND MANAGEMENT

USAIS	ISU	LOUISVILLE
EXPERIENCED YES LOW LEVEL COGNITIVE BEHAVIORS EASILY OPERATIONALIZED KNOWLEDGE FOR ITS OWN SAKE STOUTLY DEFENDED MANAGEMENT CHALLENGE USEFULNESS OF SUBJECT MATTER	EXPERIENCED YES HIGHER ORDER COMPETENCIES NOT SPECIFIED IN SELFPACED INSTRUCTIONAL PACKAGES MANAGEMENT ANALYZE AND UPGRADE	EXPERIENCED IN REVERSE TEACHERS OPERATIONALIZED BOTH STUDENT AND SELF OBJECTIVES MANAGEMENT WORKSHOP FACULTY SELF DEVELOPMENT


HIERARCHY OF MENTAL REQUIREMENTS
FOR PERFORMANCE OF JOB TASKS
HIGH
JOB PERFORMANCE
LOW
SKILLS PERFORMANCE MEASURES
EVALUATION
SYNTHESIS
ANALYSIS
APPLICATION
COMPREHENSION
KNOWLEDGE
LOW
JOB SKILLS
HIGH

SLIDE 6

REASON 2.

" PRESPECIFICATION OF EXPLICIT GOALS PREVENTS THE TEACHER FROM TAKING ADVANTAGE OF INSTRUCTIONAL OPPORTUNITIES UNEXPECTEDLY OCCURRING IN THE CLASSROOM. "

REASON 2: "UNEXPECTED OPPORTUNITIES"

EXPERIENCE AND MANAGEMENT

USAIS	ISU	LOUISVILLE
EXPERIENCED IN PASSING	EXPERIENCED YES	EXPERIENCED IN REVERSE
STUDENTS DEMAND PRECISE GOAL STATEMENTS	COURSE WRITERS (INSTRUCTORS) RESTRICTED BY PRECISELY STATED OBJECTIVES	OBJECTIVES CENTERED ON GROUP PROCESS AND STUDENT CENTERED LEARNING: NEGLECTED STRUCTURE AND ADULT GUIDANCE
ADULT STUDENTS DISCIPLINE THEMSELVES AND INSTRUCTORS WHO STRAY FROM GOALS		
MANAGEMENT REQUIRE CRITERION TESTS FOR JOB PERFORMANCE	MANAGEMENT CONTINUE TO REQUIRE PRECISE WRITING	MANAGEMENT FACULTY SELF MANAGEMENT

BEST AVAILABLE COPY

REASON 3.

" BESIDES PUPIL BEHAVIOR CHANGES, THERE ARE OTHER TYPES OF EDUCATIONAL OUTCOMES WHICH ARE IMPORTANT, SUCH AS CHANGES IN PARENTAL ATTITUDES, THE PROFESSIONAL STAFF, COMMUNITY VALUES, ETC. "

REASON 3: "ATTITUDE CHANGE"

EXPERIENCE AND MANAGEMENT

USAIS	ISU	LOUISVILLE
EXPERIENCE YES CERTAIN PROFESSIONAL STAFFERS FAVORED STATUS QUO MANAGEMENT CHALLENGE PROFESSIONALISM	EXPERIENCED YES VARIOUS LEVELS OF PROFESSIONAL STAFF ENTHUSIASM TO CHANGE IN TRADITIONAL TEACHER ROLES MANAGEMENT CHALLENGE PROFESSIONALISM ACTIVE INSERVICE PROGRAM FOR GRADUAL CHANGE OVER	EXPERIENCED YES (1) SOME TEACHERS & PRINCIPALS OPPOSED CHANGE TO "SYSTEMS" CONCEPT (2) TEACHER PARAPROFESSIONAL ROLE CONFLICTS MANAGEMENT TRANSFER OUT WITHOUT PREJUDICE

BEST AVAILABLE COPY

REASON 4.

" MEASURABILITY IMPLIES BEHAVIOR WHICH CAN BE OBJECTIVELY, MECHANISTICALLY MEASURED, HENCE THERE MUST BE SOMETHING DEHUMANIZING ABOUT THE APPROACH. "

REASON 4 DEHUMANIZING

EXPERIENCE AND MANAGEMENT

USAIS	ISU	LOUISVILLE
EXPERIENCED YES SCHOOL COURSES AND JOBS STRESS HUMAN INTERDEPENDENCE. CONTRIVED TESTS RESTRICT MEASURABLE STUDENT BEHAVIOR MANAGEMENT DEVELOP TESTS OF APPLICATION OF JUDGMENT IN JOB SITUATIONS REDUCE RELIANCE ON MULTIPLE CHOICE TESTS	EXPERIENCED YES COMPETENCIES INFERRED BUT NOT MEASURED BY PAPER AND PENCIL TESTS. INCOMPETENT STUDENTS GET BY ON NORM REFERENCED TESTS MANAGEMENT TEST OVERT BEHAVIOR REQUIRING SKILL IN PLANNING FOR, ANALYZING INTERPRETING, AND EVALUATING SITUATIONS OR BEHAVIORS ELIMINATE CONTRIVED MULTIPLE CHOICE TESTS DEVELOP TRANSITIONAL NORM TO CRITERION TEST STRATEGIES	EXPERIENCED IN REVERSE WORKSHOPS ON NEW HUMANIZED BEHAVIORAL SYSTEM" LEFT TEACHERS UNPREPARED FOR PUPIL BEHAVIORAL PROBLEMS MANAGEMENT SELF MANAGEMENT

BEST AVAILABLE COPY

REASON 5.

"IT IS SOMEHOW UNDEMOCRATIC TO PLAN IN ADVANCE PRECISELY HOW THE LEARNER SHOULD BEHAVE AFTER INSTRUCTION."

REASON 5: "UNDEMOCRATIC

EXPERIENCE AND MANAGEMENT

USAIS	ISU	LOUISVILLE
EXPERIENCED NO STUDENTS DEMAND TO KNOW IN ADVANCE THE INTERMEDIATE AND TERMINAL BEHAVIORS EXPECTED STUDENTS DEMAND OPPORTUNITY TO DEMONSTRATE PROFICIENCY MANAGEMENT DEVELOP TESTS IN A PERFORMANCE MODE OPERATE QUALITY CONTROL MODEL (SLIDE 1)	EXPERIENCED NO NOT A STUDENT COMPLAINT ABOUT SELF STUDY PROGRAMS	EXPERIENCED NO PARENTS, TEACHERS, PARA-PROFESSIONALS, AND PUPILS ARE GOVERNING BOARD MANAGEMENT AGREEMENT IN ADVANCE: PARENTS AND PUPILS ARE FINAL AUTHORITY ON NEED FOR PROJECT MODIFICATION

REASON 6.

"THAT ISN'T REALLY THE WAY TEACHING IS; TEACHERS RARELY SPECIFY THEIR GOALS IN TERMS OF MEASURABLE LEARNER BEHAVIORS; SO LET'S SET REALISTIC EXPECTATIONS OF TEACHERS."

SLIDE 15

REASON 6 "REALISM"

EXPERIENCE AND MANAGEMENT

USAIS	[illegible]	[illegible]SVILLE
EXPERIENCED YES	EXPERIENCED NO	EXPERIENCED PERHAPS
INSTRUCTORS EMULATE THEIR COLLEGE PROFESSORS IN THE KNOWLEDGE ACQUISITION WORLD	USE OF SELF PACED INSTRUCTIONAL PACKAGES IN TEACHER EDUCATION CONSTITUTES THE REAL WORLD FOR THE STUDENT IN TEACHER PREPARATION	
INSTRUCTORS RELUCTANT TO ENTER THE WORLD OF JOB PERFORMANCE TESTING		
MANAGEMENT SUCCESSIVE APPROXIMATIONS SUCCESSIVELY MORE PRECISE STATEMENTS OF JOB TEST ACTIONS CONDITIONS AND STANDARDS OF PERFORMANCE	MANAGEMENT GRADUATES EMULATE COLLEGE PROFESSORS... BECOME DEVELOPERS OF SELF PACED PROFESSIONAL SEQUENCE GUIDES"	MANAGEMENT TRANSFER WITHOUT PREJUDICE

REASON 7.

" IN CERTAIN SUBJECT AREAS, e.g., FINE ARTS AND THE HUMANITIES, IT IS MORE DIFFICULT TO IDENTIFY MEASURABLE PUPIL BEHAVIORS. "

REASON 7 "HUMANITIES"

EXPERIENCE AND MANAGEMENT

USAIS	ISU	LOUISVILLE
EXPERIENCED YES "SYSTEMS ENGINEERING OK FOR OTHERS	EXPERIENCED YES CASE IN POINT NO BEHAVIORAL OBJECTIVES FOR FLEXIBILITY	EXPERIENCED YES "AS OFTEN AS NOT", ACADEMIC OBJECTIVES WERE STATED IN MEASURABLE TERMS
MANAGEMENT SOCRATIC EXCHANGES BETWEEN INSTRUCTOR AND QUALITY CONTROLLER. INSTRUCTOR DEVELOPED HIS OWN PERFORMANCE TEST	MANAGEMENT EXPEND MORE EFFORT	MANAGEMENT FACULTY SELF MATURATION. EXPERIENCE.

REASON 8.

" WHILE LOOSE GENERAL STATEMENTS OF OBJECTIVES MAY APPEAR WORTHWHILE TO AN OUTSIDER, IF MOST EDUCATIONAL GOALS WERE STATED PRECISELY, THEY WOULD BE REVEALED AS GENERALLY INNOCUOUS. "

REASON 8: INNOCUOUS OBJECTIVES

EXPERIENCE AND MANAGEMENT

USAIS	ISU	LOUISVILLE
EXPERIENCED: YES	EXPERIENCED: PERHAPS	EXPERIENCED: PROBABLY
GOALS OF SUBJECT MATTER EXPERTS WERE EXTRANEOUS TO NEEDS OF COURSE GRADUATES	HIGHER ORDER COMPETENCIES NOT SPECIFIED	SOME BEHAVIORAL OBJECTIVES "DID LITTLE BUT PARROT POINTS OF VIEW STRESSED BY ADMINISTRATORS OR TRAINERS"
MANAGEMENT: TIGHTEN SYSTEMS ENGINEERING DEVELOP JOB RELATED TESTS	MANAGEMENT: ANALYZE AND UPGRADE	MANAGEMENT: FACULTY SELF DEVELOPMENT

REASON 9.

"MEASURABILITY IMPLIES ACCOUNTABILITY; TEACHERS MIGHT BE JUDGED ON THEIR ABILITY TO PRODUCE RESULTS IN LEARNERS RATHER THAN ON THE MANY BASES NOW USED AS INDICES OF COMPETENCE."

REASON 9: ACCOUNTABILITY

EXPERIENCE AND MANAGEMENT

USAIS	ISU	LOUISVILLE
EXPERIENCED IN REVERSE	EXPERIENCED ACCOUNTABILITY FRAGMENTED	EXPERIENCED NOT A PROBLEM
INSTRUCTORS WANTED TO TEST STUDENT KNOWLEDGE LEVELS FREQUENTLY TO AVOID BLAME FOR FAILURES	SELF PACED PACKAGES ARE GROUP PRODUCTS	TEACHER WAS FACILITATOR
MANAGEMENT USE KNOWLEDGE TEST AS INSTRUCTIONAL STRATEGY. USE PERFORMANCE TEST AS GRADUATION REQUIREMENT	MANAGEMENT CONTINUE AS PLANNED	MANAGEMENT IN PARENT TEACHER PUPIL MINIBOARD GOVERNANCE. PARENTS & PUPILS ARE FINAL AUTHORITY ON NEED FOR CHANGE

REASON 10.

"IT IS FAR MORE DIFFICULT TO GENERATE SUCH PRECISE OBJECTIVES THAN TO TALK ABOUT OBJECTIVES IN OUR CUSTOMARILY VAGUE TERMS."

REASON 10: "GENERATE OBJECTIVES"

EXPERIENCE AND MANAGEMENT

USAIS	ISU	LOUISVILLE
EXPERIENCED: YES INSTRUCTORS DESIRED TO TEACH, LEAVING DOCUMENTATION AND TESTING TO OTHERS. MANAGEMENT: SUCCESSIVE APPROXIMATIONS: TEACHING AND TESTING ARE TWO ASPECTS OF THE SAME THING.	EXPERIENCED: NOT REPORTED MANAGEMENT: ALL NECESSARY COMPETENCIES INCLUDED IN INSTRUCTIONAL OBJECTIVE PACKAGES.	EXPERIENCED: YES MEMORIZATION AND RECITATION OBJECTIVES WERE MORE BEHAVIORALLY ORIENTED THAN WERE "HUMANISTIC" OBJECTIVES. MANAGEMENT: WORKSHOPS IN BEHAVIORAL OBJECTIVES. PLATEAU PHASE: RELAX STRESS ON BEHAVIORAL OBJECTIVES. INSTRUCTOR SELF MATURATION.

REASON 11.

" IN EVALUATING THE WORTH OF INSTRUCTIONAL SCHEMES IT IS OFTEN THE UNANTICIPATED RESULTS WHICH ARE REALLY IMPORTANT, BUT PRESPECIFIED GOALS MAY MAKE THE EVALUATOR INATTENTIVE TO THE UNFORESEEN. "

REASON 11: "UNFORESEEN EVENTS"

EXPERIENCE AND MANAGEMENT

USAIS	ISU	LOUISVILLE
EXPERIENCED YES PHILOSOPHICAL CLASH OVER ROLE OF 10-MINUTE QUIZ. MANAGEMENT SEPARATE QUIZZES INTO TESTS OF CONTENT AND TESTS OF JOB TASK ACCOMPLISHMENT	EXPERIENCED YES STUDENTS LACK SELF DISCIPLINE TO COMPLETE SELF-PACED PROGRAM. STUDENTS MUST ADJUST TO UNRELENTING PRESSURE FROM PROGRAM START TO FINISH. SLOW, AGONIZING ADJUSTMENT BY STAFF AND STUDENTS. MANAGEMENT HELP STUDENTS PLAN WORK SCHEDULE.	EXPERIENCED: IMBALANCE IN SPECIFYING TEACHER AND LEARNER BEHAVIORS. MANAGEMENT: TEACHER SELF-MANAGEMENT.

CRITERION REFERENCED PERFORMANCE TESTING IN COMBAT ARMS SKILLS

John F. Hayes
URS/Matrix Company

My presentation today represents a progress report on the work that URS/Matrix Company personnel have been doing in the area of performance proficiency testing. While we have been working in this area for over four years, it is specifically our current work for the Army in the measurement of combat arms skill proficiency that I want to focus on today. This work is being done under contract with the Army Research Institute. Dr. Frank Harris, Dr. Robert Root, and Major Larry Word are the ARI personnel monitoring this work. Other project personnel from URS/Matrix include Mr. Ray Griffin, Dr. Boyd Mathers, and Mr. Don Jones.

As background to this description, I would like to present a general framework of performance testing that has evolved from our work and the work of others in this area. Performance testing itself, of course, is not a new concept. It has long been recognized as one of the more important techniques in the area of training assessment. It has, however, traditionally presented problems in terms of the administrative load that it imposes on the testing function and in terms of test reliability. It takes a great deal of time, support materials, and it requires numerous personnel to execute. Still, it represents the most valid means of assessing proficiency. This is particularly true with respect to training for specific job assignments which is a major concern of the military services.

Most general training development models call for the development of criterion tests early in the training development process. Criterion tests developed from such models are based not upon the training content being developed, but upon the same job requirement information that was generated

for the training material development itself. The test development effort, then, is a parallel and concurrent but independent effort. In this way the criterion tests measure whether the student can perform criterion job functions upon completion of training, not just whether or not he learned what was in the course.

"Criterion referenced" tests as we are employing the term means more than scores measured against fixed as opposed to relative standards. By criterion referenced we also mean that skills are measured in the context of the application situation as completely as possible. By criterion referenced we mean that the tests are designed so that performance on them represents a true measure of job performance, rather than measures of the possession of skills presumed to contribute positively to job performance. In the area of Infantry and Armor combat skills, there are no acceptable job performance conditions against which to verify or validate evaluation measures. In other areas in which we have worked, electronic and mechanical repair maintenance, work samples could be easily constructed. For the Infantry and Armor soldier this was not the case. The job of the combat arms soldier is not to apply a specific set of skills on a repetitive basis. It is to recall and select required skills as needed by unique situations and integrate and apply them under conditions of stress ranging from mild to drastically severe. What we were seeking then was to approach and approximate those conditions in our evaluation of his job proficiency.

In our development of criterion referenced performance tests, several guiding principles have become common to all of our developmental efforts.

The first is that we start by defining actual job tasks. These tasks come from the analysis of the performance of job incumbents. This information can be gathered in a variety of ways. It can be done through conventional surveys of job requirements and job duties, or by observation of job incumbents in the performance of their duties, or it can be done through the "panel of experts approach". The method used has to be selected based on the best source of information available. Such methods are not a central topic for today's discussion; however, a fundamental premise of our test development approach is that through the use of such data validity is built into the test situation from the beginning. The validity of the performance test results, then, is not demonstrated by comparison with other indices such as supervisor ratings or other test score results that are more typically used for test validation. Content validity is established as a precondition so that the final tests represent absolute criteria against which any other indices of job performance can be themselves validated.

The second characteristic of our performance tests are that meaningful size tasks are selected, which require skill integration and selection from a total repertoire of application skills. In performing a job duty, it is important that the job incumbent be able to handle all of the parameters of that job duty in the context and sequence that he will be expected to perform on the job. Under our concept, performance testing is not just the isolated performance of defined skills, but would include the selection of skills to be applied at any given time based upon the parameters of the situation. For example, a technician may be perfectly

capable of repairing a carburetor, and this could be tested specifically. It is also important in our view, that he be able to recognized when the carburetor needs to be repaired. That is, he must be able to isolate the carburetor as a malfunctioning component from among the myriad of other engine components. Further, the performance of this task should be done in the context of the operational environment. To the extent possible the individual should have to locate the problem, select tools and equipment, use appropriate job documentation effectively, correct the problem, and, in general, demonstrate his ability to operate within the operational environment. This then, is what we mean by total skill integration.

The third major characteristic of the performance oriented tests is that the measures that are used for gauging successful performance are product-oriented rather than process-oriented. When a man is functioning as part of a system he has inputs and outputs the same as any other element of that system, either man or machine. The functioning of the system depends upon the quality and timeliness of the man's outputs. These characteristics can be measured as can the outputs of any other component of the system.

Some of the more compelling aspects of performance for measurement purposes are often processes and in many instances the quality of such processes is often a matter of subjective judgement on the part of a rater or evaluator. He will observe whether the performing individual did the task in accordance with prescribed procedures without questioning whether those procedures were in fact the only way of accomplishing the output.

In the design of our performance tests, we have conscientiously selected criterion measures which do not rely upon the subjective judgement of the rater. The guiding philosophy of this test development effort was that we do not care how a man accomplishes his job or his task. What we are concerned with is whether he got the job done, i.e., produced the required outputs at the proper time. In this way, the testing procedures can remain independent of the specific processes that are prescribed.

Applying these principles to testing combat skills of the Light Weapons Infantryman (MOS 11B), we began with a definition of the situations in which the individual is expected to perform. These can generally be defined in terms such as: assault and defense, reconnaissance patrol, defensive patrol, night infiltration, and anti-tank actions. The next step was to perform a behavioral task analysis, in cue-response terms, to determine exactly what the individual soldier is responding to in each of these situations (the cues) and what outputs (response) are expected of him.

One of the first characteristics that became evident was that while we were talking in terms of individual proficiency and individual testing, the soldier seldom acts as an individual in the defined criterion situations. He is virtually always a member of a larger unit, most often the fire team, operating as a member of a squad. As we proceeded, it became clear that we had to test him in that context.

Another critical factor that became clear is that the infantryman is operating most of the time against an intelligent opposing force, that is only vaguely predictable in its actions and which can kill him unless it

is neutralized. Initially this element was considered too dynamic and too difficult to provide to be considered for inclusion in a test. The more we examined the situation, however, the clearer it became that without this element, attempts to test the combat efficiency of the individual soldier were going to be woefully inadequate. While we investigated hit-indicator systems such as laser beams and radio activated equipment, we felt strongly that, to be effective and usable, any system that we devised had to be free of complicated equipment.

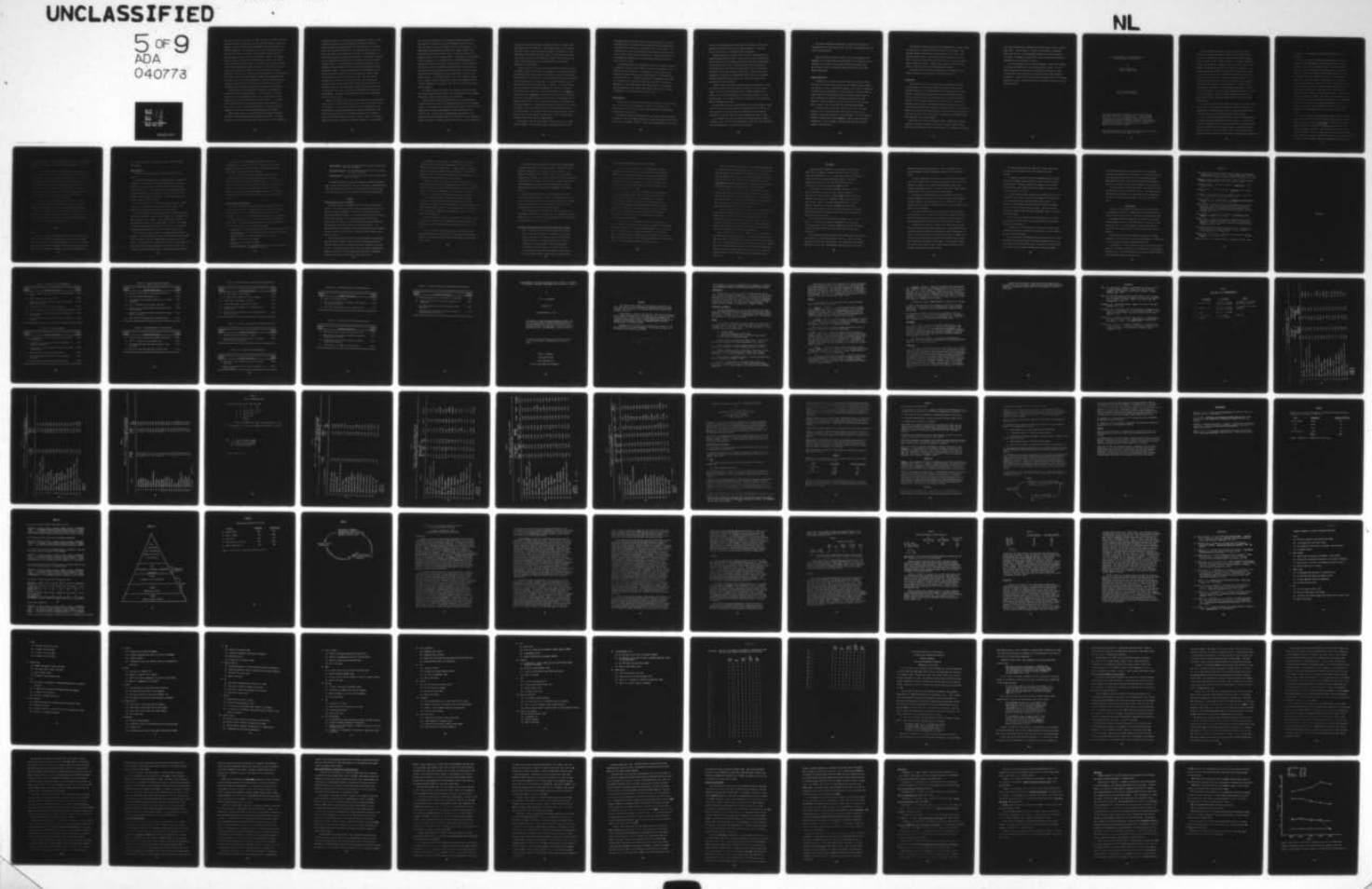
Eventually we devised a system which uses a two-digit number placed on each individual's helmet and rifles equipped with telescopes. When an infantryman can distinguish an opponent's number through his scope and fires a blank round, it is assumed that he would have been able to kill that individual. By placing test controller personnel with each of the opposing elements and equipping the controllers with a radio, individuals can be knocked out of action on each side as they are effectively engaged by the opposing force. The problem is thus allowed to play itself out on a realistic basis. By experimenting with combinations of color, background, and size of the numbers and the power of the scope, a combination of number-color and number-size, and scope power, was achieved which duplicates closely the effective range of the M16 rifle. For all practical purposes this range is 200 meters. While the actual range of the M16 is much longer, the combat situation seldom provides accessible targets beyond that distance.

The developing of this scoring system opened new possibilities to the use of combat situations for testing purposes. The situations could

AD-A040 773 AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TEX F/G 5/9
PROCEEDINGS, ANNUAL CONFERENCE OF THE MILITARY TESTING ASSOCIAT--ETC(U)
NOV 73

UNCLASSIFIED NL

5 OF 9
ADA
040773



now provide realistic cues which are highly similar to the combat situation. They also provide the opportunity for the individual soldier to select and employ, in concert with his fellow squad members, what he considers to be effective responses. The required responses are not pre-defined; he has to select them from his repertoire, under stressful conditions, at the most appropriate time, in order to be successful. Further, the situation will now change as a result of the individual acts taken on both sides. When the individual perceives that he is affecting the outcome, he becomes a very interested and enthusiastic participant in the exercise. When the situation does change as a result of actions one takes, it becomes a measurement criteria. One of the most important elements of the infantryman's life is that fact that he must constantly respond to an environment that is changing under very stressful conditions. And his responses cannot be isolated in terms of his own situations as an individual, but must be coordinated with his fellow squad members, if they are to be effective.

Techniques were then developed for simulating and measuring the effects of the other weapons available to the infantryman. The tests now permit the use of grenades, claymore mines, booby traps, anti-tank weapons, machine guns, and artillery in realistic fashion. As a result, the overwhelming opinion of experienced combat infantry personnel is that we have, in fact, created situations which call for very realistic responses in conditions very closely approximating actual combat.

Having achieved this, it was then necessary to apply the necessary controls to achieve individual testing within this context. The basic data-gathering instruments during these situations are the test controllers

previously mentioned. In addition to being equipped with radios in order to make the situations play realistically and to assess casualties, control personnel have responsibility for recording product outputs for each of the individual participants. The controller keeps track of who scored what kills or was killed with rifle fire, who threw grenades or was killed by grenades, who called in artillery or was killed by artillery fire, who employed booby traps, mines or claymores, and who was killed by these devices. The mechanisms developed for keeping track of this information rely primarily on the use of the identifying numbers on each individual's helmet. Points are accumulated for each individual on the basis of the number of kills that he has scored, plus a factor added for the degree of success achieved by this unit. The unit score is determined on the basis of a set number of points for achieving the objective, less a set number of points for each casualty incurred. The individual scores are then loaded with a factor based upon unit success. At the present time, both the defense and the offense are being scored simultaneously. Therefore, the degree of success of one is diminished by the score of the other.

There are, of course, many chance factors operating in an individual's score, just as in combat there are many chance factors operating to affect one's mortality. A man can get killed by his own squad members, he can be killed by an ill-launched grenade which happens to bounce off a tree back into his own position, or by calling artillery fire on himself. Some of these are his errors,some are the errors of others. To overcome these chance factors, repetition of testing is employed. Going through a

test situation one time will not necessarily produce an index of the best members of a squad, or absolute individual proficiency scores. Over the course of a series of repetitions of the test, however, it is hypothesized that proficient individuals will achieve higher scores than those who are not proficient. That is, the chance factors will gradually balance out for all participants. Furthermore, it is necessary for the situation to be run numerous times to provide exposure of the individual to different situations. Even running the same problem over the identical terrain produces an altered situation. The fact that no two problems ever are run alike with exactly the same results or events occurring is also assumed to be a valid construction in terms of the combat situation. Each combat situation is, in fact, unique and being able to respond properly to unique situations is the highest measure of a combat infantryman's proficiency. This then is what we mean by total skill integration.

To provide a better description of the system in operation I have some slides taken during our tryout of the Infantry tests at Fort Benning which depict several of the criterion test situations.

Testing situations were developed for each of the other defined criterion situations using the same guiding principles as for the assault/defense. The individual was to receive the same inputs, perceive the same environment and make the same response as would be required by the criterion job. If there was a non-predictable enemy present in the criterion job he was provided in the test situation. In the reconnaissance patrol, for example, a squad was given an objective to recon, a map and an

operations order describing what they might encounter in trying to gain intelligence about the objective. The defensive patrol on the other end was given a sector to patrol and protect against recon elements. They could employ any strategy they chose, such as a fixed screen with booby trap, a moving screen, or any combination thereof. The problem was then started at a prearranged time and it ended when the recon element returned or was neutralized.

Upon the return of the recon patrol all surviving members were tested by having to report what they saw at the objective. The squad was awarded points based upon the percentage of intelligence returned by any members. The objective of a recon patrol is to get accurate and complete intelligence about the enemy back to the base camp. If only one man gets back but he has complete information, the mission is a success. If all get back but with only sketchy or inaccurate information the mission is a failure. The scoring system, therefore, has to be constructed so as to promote the correct outcome and penalize incorrect ones. One of the easiest traps to fall into in performance testing is to construct a scoring system that is simple, objective, easy to use, and directly opposed to the promotion of the desired behaviors. Overloading the reward for surviving or for correct performance of enabling processes without due consideration of results, is self-defeating. The constant concern with test and training situations that are this realistic must be that they do not teach, through design or default, habits that will get people killed.

The Infantry tests were field tested utilizing troops of the 197th Infantry Brigade at Fort Benning, Georgia, during August of this year.

The experimental design plan called for the total pool of test items to be administered to a trained Infantry platoon that could be expected to know and be able to perform at a reasonable level of proficiency. The individual test items that correlated best with total test results were then to constitute the final proficiency test. Further, this tryout was to determine the feasibility of the scheduling and times allotted for each of the test items.

As indicated earlier, no independent criteria were considered suitable as validation criteria, since no other criteria were considered to be as valid as the tests themselves. The elements of the test content were reviewed by numerous infantry experts and judged to be valid samples of the infantryman's job. Therefore, agreement or disagreement with indices such as supervisors' ratings or written test scores was not considered appropriate. To determine whether the results did agree with commonly accepted judgements, peer ratings were employed which asked for ratings of preference of individuals to go into combat with.

Test Evaluation

At this point in time we have only had an opportunity to take a cursory look at the results achieved during the tryout of the Infantry tests. Following completion of the Armor test tryout more complete analyses will be conducted of all the data.

The Infantry tests were administered to one Infantry platoon consisting of 35 people. Nine were Squad and Fire Team Leaders (designated as MOS 11B.40) and 26 were squad members (11B.20). Formal test administration took one week.

Prior to test administration pre-training was conducted to insure that personnel were familiarized with the proper employment of the rifle scope, since that was a non-standard item of equipment.

Several weeks were also spent prior to the test administration in training control personnel (NCO's) in the details of test administration and in insuring that the Infantry personnel had sufficient minimum skills to participate in the tests. This was necessary since the platoon we were furnished was newly formed with personnel just out of individual training who had no unit training or experience. Since the tests were designed for journeyman level infantrymen, basic skills in map reading, artillery adjustments, squad tactics, etc. had to be covered so that these personnel could be tested.

After the week of testing the platoon members were requested to rank each other in terms of whom they would most prefer to have with them in combat. While they had participated with an against everyone in the platoon, no one had any knowledge of the test score results at the time of this rating. It was based on what the individuals had seen during the month of training and testing.

The raw scores for the 2 level personnel (n=26) (squad members) ranged from 75-227 with a mean of 214.96 and a standard deviation of 48.26. For 4 level personnel (n=9) (Squad and Fire Team Leaders), the range was 158-434 with a mean raw score of 335 and a standard deviation of 96.34.

The correlation of total test score with the peer ratings, computed by Pearson r, was .697 for 4 level personnel and .420 for 2 level personnel. Both are significant at the 05 level of confidence.

The Combat Performance sub-portion (two-sided) of the total test correlated with the total test score at .942 and .927 respectively for the 2 and 4 level personnel.

Due to the built-in content validity of the test situations, the agreement of test results with peer rating data, and the spread of scores obtained, we are satisfied that these situational performance tests are yielding an index of individual combat proficiency for the Light Weapons Infantryman.

Armor Crewman Tests

A similar set of criterion tests were developed for Armor crewmen. The duties of the Armor crew positions are well defined and they combine with the tank to make a single weapons system. This is both a simplifying and complicating characteristic. The end product of a tank gun round on target involves the coordinated actions of four people which makes the job of ferreting out individual contribution and proficiency more difficult. The proficiency of the tank crew as an entity is, and has been, a common level of distinction for proficiency measurement.

To break through to individual proficiency we went to a process checklist type of evaluation within each situation when it was not possible to get product measures on each individual. The tank criterion situations include tank-to-tank engagement, night bivouac situations, infiltration, a river crossing, overcoming mined obstacles, bunkers, ambushes, and maintenance.

The primary scoring system used in tank engagement is to sight through the main gun via a telescope as with the infantry tube weapons. This provides immediate verification to the controller as to whether a hit would have been achieved. He can then take the appropriate action to kill the other tank (via radio) or notify the crew that the round missed and the situation continues. Against enemy infantry, the same scoring system is used as before that of reporting helmet numbers.

The Armor tests are currently undergoing field evaluation at Fort Carson, Colorado, and no results are available at this time.

Conclusions

Our experience to date in this project has reinforced our belief that performance testing could be employed to measure larger and more meaningful segments of job performance than has generally been the case to date. Certainly the control and data collection problems increase, but the increased validity of the exercise and the resulting increase in confidence of the results appears to be worth the costs. The change in emphasis is from having the individual demonstrate that he knows or can perform each individual skill required, to having him recall and effectively apply such skills as the situation demands. While there may be less control over which specific skills are tested, the sampling can be handled through both repetition and situation design.

These tests are not yet in the final format. Refinement and further research is required to improve the quality of the data collected and to increase the discrimination power of the results. We strongly feel, however,

that these are problems of technique and not the inherent nature of performance tests. The challenge is to pursue the refinement of techniques rather than back off to more manageable though less fruitful approaches to testing. Performance testing has the power of accurate job proficiency assessment and the challenge is to tap it.

As to the increased costs in time and manpower, we feel that a change in attitude and commitment to the testing process is needed. If we now have the beginning of mechanisms for accurately determining the status of proficiency of our personnel, the resources should be made available to employ it. Without the type of data that is available from well designed performance testing, training efforts will necessarily operate at less than optimum efficiency.

THE AIR FORCE WIFE: HER KNOWLEDGE OF, AND ATTITUDES TOWARD, THE AIR FORCE*

By

John A. Belt, Ph.D &
Arthur B. Sweney, Ph.D.

Center for Human Appraisal
Wichita State University

This paper prepared for presentation at the Military Testing Association Conference on "Human Resources - Growing Demands," Oct. 28 - Nov. 2, 1973, San Antonio, Texas. A more analytic treatment of the material presented in the paper may be found in technical report #103, "The Air Force Wife - A Study of Morale among Military Dependents" issued by the Center for Human Appraisal, Wichita State University.

*The Research reported in this paper was supported by the Air Force Office of Scientific Research, grant # 72-2001.

A major corporation gives gold charms in the shape of each new state to wives of transferred executives and it becomes status to point with pride at the number of charms acquired during a career span. A novelty? Public relations gimmick? No, rather a small part of the concerted effort in industry to make wives feel a part of their husband's career and satisfied with their place in this future. The industry motives are far from altruistic for as William Whyte stated in 1948 "as an econoic lever...companies have learned that there is no stimulus quite so effective as the wife if properly handled." Much research has gone into ways to improve the recruiting and retention of civian employees and maximize their productivity. Helfrich, (1965) concluded that "corporations are increasingly interested in the wives of their executives." In an 1971 study of business executives J.M. and R.F. Pahl typified the comments of their interviewees on the single most important factor influencing their career, "my wife, more than anything else."

It would logically follow then that the military in many ways would not only follow this pattern but demand more from wives and family. And likewise, the commitment a man makes in choosing a military career would call for support and approval from his wife and family.

As our society has become more and more mobile, the extended family of the past has been replaced by a more insular unit. No longer does a man have relatives close by or living in his home, for his many moves may take his family far from any familial ties. Without these other sources to rely on, the family tends to turn to each other and inter-personal relationships attain a higher value in assessing each individual's

satisfaction. The demands of a military life emphasizes this phenomenon.

One wife in a recent survey of military families related that her eldest child had attended 14 schools before entering college. (Entrails, 1971) How she views this style of life, the rewards it offers her and her family must be weighed against the deficiencies. And the conclusions she reaches concerning these matters will play an important role in her husband's life also. He has drawn his family into a life style unique in many aspects; a subculture of over 4 million dependents. Roger Little (1971) has remarked on this, "All military families have in common knowledge and experience in an occupational culture (or subculture) which is more distinct than that of other occupations in the larger society." Base housing, required mobility, the status in the community of the military man - these are only a few of the pressures a man choosing a military career must weigh. Add to these the similar pressures his wife and family meet and one can see either an effective cohesive unit or a point of dissention for a man trying to fulfill a role as a husband and father.

In a recent survey of career attitudes among Air Force personnel, it was found that the wife had twice the influence upon her husband's career intentions as any other individual, including his immediate supervisor or any of his peers.† (Belt, 1972) A determinant as important as this can't be slighted or ignored. In an investigation of junior officer retention problems, Lund (1972) found that wives were the key variable in the decision to separate or remain in the army. A 1971 study emphasized that any good junior officer retention program must include efforts

aimed toward the wife. "The military family has come to be considered an essential component in personnel policy and management." (Lund, 1971)

How a man's family perceives him and his role in the community exerts tremendous impact on his own view of himself and the value of his work. This affects not only his attitude but increasingly his performance and his relationships with those around him. Much effort has gone into studying the military man as an important resource in our society today, particularly with the changes being affected now, but these studies would indicate that one of the most crucial variables in attaining full value of the individual to society has been at least qualitatively ignored.

To understand the role a military wife plays in her husband's life, one must first analyse her role in the military and then expand this to examine how her position as wife and her attitudes as a member of the military (wives in responding to questions tend to answer "when we were stationed at McConnell" or "we have been in the service 7 years") relate and exert influence within the family. This study was undertaken for just such purposes.

METHOD

Subjects

The entire sample was drawn from the wives of men assigned to the 381st Strategic Missile Wing, stationed at McConnell Air Force Base. The wing consisted of two operations squadrons, one missile maintenance squadron, a communications squadron, a security police squadron, and various support squadrons and groups. There were approximately 1600

men in the wing during the time of the study, over half of which were married.

Data collection

A series of three questionnaires were administered to the subjects.

The first questionnaire (SDS-1) was developed as an exploratory instrument to investigate the basic premise that wives of Air Force personnel do, in fact, have strong opinions about their husband's military affliation. SDS-1 contained questions from major catagories: Demography; Relations with Facilities, Benefits, and Services; and General Attitudes toward Military Life. An open ended section for general comments was also included.

As a pilot study, SDS-1 was distributed to the Hospital Squadron to check on its applicability to the Air Force life style. Upon analysis, the instrument was determined to be applicable.

The first questionnaire was administered by distributing the forms to all the married men in the wing, and requesting the husbands to take the instruments to their wives. Return envelopes were provided to facilitate the return of the completed forms. A total of 827 copies of SDS-1 were distributed on April 14, 15, 16, 1971. Of these, 264 completed forms were returned, creating a response rate of 31.9%.

Wives Attitude Survey II (WAS II) was generated for two major purposes. The first section was aimed at investigating in depth trends that were evident in the open ended response section of the first survey. The second aim was to measure more precisely the relations the wives had with the facilities, benefits, and services. Again, the instrument included a demographic section and an open response section.

A listing of all home addresses were obtained through the Consolidated Base Personnel Office. The second survey was mailed directly to the wives on March 13, 1972. Included in the mailing were stamped, self addressed return envelopes.

Wive's Attitude Survey III was created primarily to explore in more depth the relation between the wife and the facilities, services, and benefits provided by the Air Force. This investigation included further inquiry into the source of the wives' knowledge and the wives' perceived need for the Air Force to offer these items. Demographic information and general comment sections were included. The questionnaire was distributed on July 18, 1972 in the same manner as WAS II.

Statistical Methods Employed

Due to the disparate nature of the data, two separate techniques for analysis were used. A factor analysis was performed on variables 1-25 and 50-55 of WAS II to determine the dynamics at work in the attitudes of the wives. A discriminate function analysis was performed in a stepwise manner across pairs of ten selected groups and subgroups of the sample. These groups were operational defined by reflection of their husbands service in the Air Force as follows:

Missile Crew - men that are assigned to operational Combat Missile Crews.

Non Missile Crew - men that are not assigned to Combat Missile Crews.

Career - men with five years or more of time in service.

Officer - men of rank O-1 or above.

Enlisted - men of rank E-9 or below.

First Term Officer - men with four years or less of time in service with rank O-1 or above.

Career Officer - men with five years or more of time in service with rank of O-1 or above.

First Term Enlisted - men with four years or less of time in service with rank E-9 or below.

Career Enlisted - men with five years or more of time in service with rank E-9 or below.

The dyads selected for analysis to determine intergroup differences were: Missile Crew-Non Missile Crew; First Term-Career; Officer-Enlisted; First Term Officer-Career Officer; First Term Enlisted-Career Enlisted; First Term Officer-First Term Enlisted; Career Officer-Career Enlisted.

RESULTS

Factor Analysis of Wives' Attitudes

The Factor Analysis performed on section of WAS II generated ten factors. The contributing variables and their loadings for each of the factors are shown in Appendix 1. The first factor was described as passive alientation/integration. It canoted a passive role for the wives of Air Force personnel. They apparently did not feel that their participation was required or even solicited, yet neither did they feel that they were rejected or prevented from becoming involved. This dynamic displayed the disavowal of personal responsibility for integration into the Air Force lifestyle and an attitude of simply "floating along with the current."

Factor II was identified as a desire for information versus apathy. This continum was between a desire for more information about how the Air Force affects her life and the apathy which is present in all walks of life. These feelings were active in that the dynamic stretches from apathy to curiosity. It also appeared that there was a recognizable solidarity of interest in what the Air Force was and did.

The dynamic at work in Factor III could most aptly have been labeled familial maturity and independence from the Air Force/ familial immaturity and dependence on the Air Force. The dynamic was one of growth and change in focus of attention. As the family matured, the wife became more interested in its development and less interested in her relationship to the Air Force.

Factor IV provided an insight into the wives' perception of the current societal trend of distaste for the military. The continum traversed the area between prideful identification with her spouses' job and the couple's relationship to the Air Force to apologetic recognition and rejection of identification with the Air Force.

The description of Factor V was a facet of wives' attitude that was not directly related to the Air Force as a functioning organization. Instead, it was an indication of the group identification/ disassociation among the wives of Air Force personnel. This within group bipolarity was apparently a very important reality of life in the Air Force community.

The dynamic exposed in Factor VI was difficult to interpret. Without referring to causality, it appeared that physical/psychological separation or proximity/identification attitudes were present in the sample.

Factor VII probably revealed an experential attitude set about regulation/restriction by the Air Force. The dynamic between wives' feelings of personal freedom of action and restriction of action was quite obvious.

The eighth factor was interpreted as two varying perceptions of the adage "rank hath its priviledges (power)". It was based on the perceived transference of the husband's rank to the wife. The dynamic was primarily one of perceived power in the transference as opposed to no power. The younger, lower rank, wives felt more power was inherent in the transfer, while the older, higher ranking wives were frustrated by the erosion of their preconception with the realization of the lack of power that the husbands' rank gives to his wife.

Identification with the source of information about the Air Force was the interpretation of Factor IX. It became apparent that as the wife learned more about the system with which she is involved, she identified more closely with the source. The poles of this continum were the Air Force itself and her spouse.

The last factor was viewed as the conscious commitment of the wife toward involvement or noninvolvement. The frustration dimension did not enter into this dynamic, as participation was directly related to the personal decision of the wife.

Discriminant Function Analysis of Selected Groups and Subgroups

(Note: The nature of the data array required that separate discriminant function analysis be performed on each of the three instruments. The voluminous nature of the resulting analysis (21 pages of tables) precludes any but a cursory treatment of the findings here. The interested reader is referred to the tech report #103 "The Air Force Wife - A Study of Morale among Military Dependents" issued by the Center for Human Appraisal, Wichita State University for a

detailed and comprehensive analysis of this data.)

Missile Crew-Non Missile Crew. - The group of wives whose husbands were in Combat Missile Crews were discriminably different than the rest of the test population. The main factor underlying this difference appeared to be a lack of association between the wife and the Air Force, emphasized by a negative attitude on the part of the Missile Crew wives toward facilities, benefits, and services offered by the Air Force.

First Term-Career. - The many obtained differences between the First Term and Career groups could be expected because of the obvious age and time in the service differentials. The first tendancy was for the First Term wives to be more immature than their older counterparts, with the difference showing up in varied perceptions of facilities and services offered to them. The Career wife, since she had more time in the service, was more familiar with the benefits that were available, and maintained a more positive attitude toward them.

Officer-Enlisted. - In comparison to the Enlisted wives, the wives of officers were basically more socially orineted and participated to a greater degree in the Air Force centered community. In conjunction with this participation, the Officer Wife was more cognizant of the benefits available on-base, and was much more selective in the use she made of Air Force offerings. These inclinations result in the observation that the Officer wife identified more with the Air Force.

First Term Officer-Career Officer. - In the subgroup comparison of First Term Officer wives and Career Officer wives the results were that the career wives identified more closely with the Air Force, knew more about it, and generally maintained a stronger relationship with the Air Force. The findings that the career wives were more knowledgeable was further emphasized by the fact that the First Term Officer group was more influencable by military propaganda.

First Term Enlisted-Career Enlisted. - Though the First Term Enlisted and Career Enlisted wives appeared to differ in primarily the same ways as the First Term and Career Officers, the distinction was slightly different. The First Term Enlisted had a more negative opinion of the Air Force, along with the expected patterns of disassociation and ignorance of the Air Force.

First Term Officer - First Term Enlisted. - The differentiation shown between the First Term Officer and the First Term Enlisted drew along the same line as the overall rank differential. The Officers wives were more associated with the Air Force, though there was still a significant thread of disinterest in the discrimination. Again, greater maturity was noticable in the wives of the First Term Officers, and larger self perception of independence.

Career Officer-Career Enlisted. - Again, the difference between the Career Officer's wife and the Career Enlisted wife appears to have been one of rank. The Officer wife felt more socially interested in the Air Force, and was generally more associated with the Air Force. Further, the Officer wife appeared to make more conscious decisions about which facilites, benefits, and services were beneficial to her, and did not appear positively inclined toward something merely because it was 'Air Force.'

DISCUSSION

Although numerous complaints were voiced, the wives for the most part had a favorable attitude toward the military lifestyle and their participation in it. In fact, most of the wives expressed a desire for more information about the Air Force. The study indicated the average AF wife knows very little about the facilities, benefits and services available to her.

Nevertheless, the dependent briefings, designed to offer this type of information, seemed to be an irritant to some wives, especially the younger group. A fairly common complaint about the dependent briefings was the "impersonal" manner in which the wives were treated. Many also disliked the term "briefing." The term seemed to accentuate the military atmosphere of the meetings. Because of this attitude, many of the wives said they did not attend the dependent briefings regularly, which forced them to look for an alternate information source.

This alternate source usually was the husband. But he too was often an insufficient source for information. For one thing, the husband doesn't know the kinds of things his wife wants and needs to know and doesn't appear to be interested enough in these areas to find out.

The information factor should be of major importance to AF officials as the survey also indicated the wives, as a group, tend to identify and form attitudes about the AF based on the information they received. However, the survey also indicated that coercion would NOT be a good

means of attaining the desired goals. That is, requiring wives to attend dependent briefings, undoubtedly would have a very negative effect.

As might be expected the wives of men with long term associations with the AF tended to have a prideful identification with the Air Force. Some of the younger wives, however, seemed apologetic about their husband's military affiliation. This may be due to the recent negative societal view of the military among some young people.

The survey indicated that the wives who identified with the AF lifestyle tended to live close to the base, and those least interested in that lifestyle lived further from the base. However, it is not possible from this study to determine which is the cause or which is the result of this phenomenon.

It was also found that wives with no children tended to depend on the military to give structure to their lives. However, as children came into the family unit, the wife changed her viewpoint and became more involved with her family and less involved with the military.

An interesting phenomenon revealed by the survey was the concept some of the wives had about rank transference. It appeared that wives of lower ranking men perceived a great deal of power transferred to the wives of higher ranking men (both officer and enlisted.) However, when the men rose into the higher ranks, the wives realized there was very little real power transferred to the spouse. The frustration seemed to increase as the husband's rank increased and the realization that she had no real power became more evident.

The various group comparisons made in the study revealed that the feelings and attitudes of the wives tended to be group specific.

For example, the group of wives with husbands in missile crews had the most negative feelings about the Air Force and its benefits and facilities. This was not surprising, however, as our previous Career Attitude Survey showed missile crew members to be among the least satisfied men in the AF.

One factor that may have contributed to the poor attitude of the "missileer's" wife was the fact that her husband's duty required him to be separated from her overnight several nights a week. This factor may be compounded by the fact that most missile crew members were first-termers and were relatively young. The wife's younger age may have tended to make her less understanding about her husband's reoccuring absence.

Wives of the career group men were more familiar with the benefits available to them and had a more positive attitude toward them than the first term-wives.

In comparison to the enlisted wives, the wives of officers were basically more socially oriented and participated to a greater degree in the AF centered community.

The career officer wives identified more closely with the AF, knew more about it, and generally had a stronger relationship with the Air Force than the first-term officer wives. The first term officer group also seemed to be more readily influenced by military propaganda.

The difference between the career officer wife and the career enlisted wife appeared to be one of rank. The career officer wife felt more socially interested in the AF and generally was more associated with the AF. Also, the officer wife did not appear to approve of a program or benefit merely because it was 'Air Force.'

By dividing the wives into four groups it was possible to rank them according to their feelings about the AF. The career officer wife was the most favorably inclined towards the AF, followed by the career enlisted, first term officer, and first term enlisted wife.

IMPLICATIONS

Overall, it would appear that, because of her obvious effect on the husband, the AF wife's somewhat positive feelings about the Air Force and her desire for more information are desirable effects.

Her apparent desire to become more of a part of the military centered community seems to be blocked because of a lack of an easy aveneue to do so. The results seemed to indicate that both she and the Air Force lacked the real initiative to draw her into the community.

The apparent failure of the dependent briefings in providing information to the wives should be given some attention. It seems that a more "feministic" approach is desired by the wives.

In conclusion, it seems evident that the AF, as well as the other services, is ignoring a segment of its community that is an extremely powerful factor in retention and possibly an equally powerful influence on job performance and morale among the men.

REFERENCES

Belt, John A., and Parrott, Gerald S. The Relationship of Satisfiers-Dissatisfiers In A Military Unit to Re-enlistment, paper presented at Inter-University Seminar, Chicago, Illinois, September 21,-22, 1972.

Entrails, research newsletter published by Center for Human Appraisal, Wichita State University, Vol. 1, No. 2, Nov. 1, 1971.

"Eternal Triangle ... Man, Wife, and Work". Industry Week, 168:4. January 25, 1972.

Farber, Leslie E. "He Said, She Said". Commentary, 123:24. March, 1972.

Helfrich, M.L. The Social Role of the Executive's Wife, Bureau of Business Research, Ohio State University, 1965.

Little, Roger W. "The Military Family". Handbook of Military Institutions, 247-272. Beverly Hills: Sage, 1971.

Lund, Donald A. "Problems of Junior Officer Retention in the Volunteer Army: The Case of the Military District of Washington". Paper presented at the Workshop on Military Manpower - The All Volunteer Military. Inter-University Seminar, Chicago, Illinois September 21-23, 1972.

Lund, Donald A. "Active Duty - Yes or No?" The Military Police Journal 20 (February): 15-17, 1971.

McKain, Jerry L. "Feeling of Alienation, Geographical Mobility, and Army Family Problems: An Extention of Theory". Published Dissertation: Catholic University of America. National Technical Information Service, 1969.

"Navy Wives" Perceptions of Conditions of Navy Life". Naval Personnel Research and Development Laboratory. Defense Documents Center, Washington. D.C. March, 1971.

Triebal, Joane. "Your Wife: A Prisoner of Your Success?" Nation's Business, 73:13, June 26, 1972.

Whyte, William H., Is Anybody Listening? Doubleday, New York, 1948.

APPENDIX 1

Factor I: Passive alienation/integration

Variable Number	Variable Description	Factor Loading
4.	The A.F. doesn't care what the wive's of its personnel think.	-0.889
13.	The A.F. doesn't care what I think.	-0.884
7.	There are no procedures for me to express my feeling about A.F. policies.	-0.788
10.	Civilians don't respect military personnel.	-0.570
19.	The A.F. keeps the wives of its personnel well informed.	0.559
24.	Life as an A.F. wife provides me many opportunities to get involved.	0.402
5.	A.F. wives should be kept better informed of base activities.	-0.369

****

Factor II: Curiosity/apathy

Variable Number	Variable Description	Factor Loading
15.	If I understood it more, I think the Air Force would be interesting.	0.739
14.	I would like to be invited to attend my husbands re-enlistment interviews.	0.707
9.	I don't want to know more about my husband's job than I already know.	-0.701
8.	Base activities have a direct affect on me.	0.526
5.	A.F. wives should be kept better informed of base activities.	0.374
11.	Most wives think they have their husband's rank.	0.322

****

Factor III: Familial maturity/immaturity

Variable Number	Variable Description	Factor Loading
51.	How long have you been in the service (4-8yrs.:5)	0.860
53.	How many children do you have? (1.2)	0.788
52.	Do you expect your husband to make the A.F. a career (probably)	-0.727
22.	The A.F. should not be a 24-hour a day job.	0.590
24.	Life as an A.F. wife provided me many opportunities to get involved.	-0.480
1.	Wives of A.F. personnel should be involved in formulating A.F. policy.	0.360
11.	Most wives think they have their husband's rank.	0.305

****

Factor IV: Prideful/apologetic identification

Variable Number	Variable Description	Factor Loading
20.	A wife should be proud of her husband's profession.	-0.889
25.	The A.F. should not be just another job.	-0.553
5.	A.F. wives should be kept better informed of base activities.	-0.413
11.	Most wives think they have their husband's rank.	-0.340

****

Factor V: Within subculture identification/disassociation

Variable Number	Variable Description	Factor Loading
18.	A.F. wives have a lot in common.	0.825
12.	A.F. wives have a number of similiar problems.	0.722
17.	I often feel I am a member of the A.F.	0.702
2.	I enjoy associating with other A.F. wives.	0.565
21.	I know the wives of members of my husband's unit fairly well.	0.540
24.	Life as an A.F. wive provides my many opportunities to get involved.	0.474

****

Factor VI: Physical & psychological proximity/separation

Variable Number	Variable Description	Factor Loading
54.	How far do you live from base? (3-5mi)	0.885
1.	Wives of Air Force personnel should be included in formulating Air Force policy.	-0.370

****

Factor VII: External restriction/freedom

Variable Number	Variable Description	Factor Loading
6.	Wives of Air Force personnel are free to do what they want.	0.790
3.	Many of my personal activities are regulated by Air Force policies.	-0.498

****

Factor VIII: Rank transference aspiration/frustration

Variable Number	Variable Description	Factor Loading
55.	What is your husband's rank.	0.832
24.	Life as an Air Force wife provides me many opportunities to get involved.	-0.396
11.	Most wives think they have their husband's rank.	0.367

****

Factor IX: Identification with information source

Variable Number	Variable Description	Factor Loading
26.	Where did you learn the most about the Air Force.	0.880
1.	Wives of Air Force personnel should be involved in formulating A.F. policy.	0.270
14.	I would like to be invited to attend my husband's reenlistment interview.	0.268
17.	I often feel I am a member of the Air Force.	-0.250

****

Factor X: Personal latitude for involvement/non-involvement

Variable Number	Variable Description	Factor Loading
23.	Air Force benefits do not interest me a great deal.	0.684
16.	I don't care about Air Force policy except as it affects me.	0.581
5.	Air Force should be kept better informed of base activities.	-0.442
1.	Wives of Air Force personnel should be involved in formulating Air Force policy.	-0.392

****

CADET RETENTION AT THE ROYAL MILITARY COLLEGE OF CANADA AS A FUNCTION OF PERSONALITY, LEADERSHIP PERFORMANCE, AND BIOGRAPHICAL VARIABLES

by

LCol G.J. Carpenter

October 1973

Research Report No 73-4

The research for this paper was supported in part by the Defence Research Board of Canada, Grant Number 9465-12. The views expressed are those of the author and do not necessarily reflect official views or policy of the Royal Military College or the Department of National Defence.

Delivered 31 October 1973 to the 15th Annual Conference of the Military Testing Association at San Antonio, Texas.

Gerald J. Carpenter

Lieutenant-Colonel

Head, Department of

Military Leadership and Management

Here's one

ABSTRACT

This study concerns prediction of retention in training at an institution where voluntary departures are more frequent than academic failures, and the prediction of voluntary attrition has proved to be extremely difficult.

Using the cadet intakes of 1970 and 1971, data from two personality tests, leader performance scores from cadet wing peer evaluation reports, and biographical information from a 92-item Personal Preference and Biographical Questionnaire (PPBQ), several analyses were carried out to ascertain those variables which are of greatest value in the early identification of voluntary leavers.

Predictions for the first year of training of the intake of 1972 were made based on information gathered in the previous two years, and the outcomes are reported and discussed.

I'm working on mine right now!

N.

CADET RETENTION AT THE ROYAL MILITARY COLLEGE OF CANADA AS A FUNCTION OF PERSONALITY, LEADERSHIP PERFORMANCE, AND BIOGRAPHICAL VARIABLES

INTRODUCTION

My aim today is to tell you about the Royal Military College (RMC) of Canada at Kingston, Ontario, and to describe a research project which studied cadet attrition. We have found that academic failure is relatively easy to predict and that selection standards can be adjusted accordingly. However, at the moment, the rate of voluntary withdrawal from the college is about four times that of academic failure. The present study, therefore, is concerned with voluntary leavers.

BACKGROUND INFORMATION

RMC was opened nearly 100 years ago, was patterned after West Point, and offers 4-year university programmes in Engineering, Science, and Arts. The college also has a graduate school of Engineering and of War Studies. The cadet population is approximately 600. The programme is both physically and mentally demanding.

METHOD

The following instruments were used to gather data with which to test the hypothesis that cadets who withdraw from the college are not different from those who stay. (Table 1 provides information on time of testing and retesting if applicable.)

1. Cattell's 16PF.
2. Jackson's Personality Research Form.
3. RMC Personal Preference and Biographical Questionnaire.
4. RMC Cadet Peer Rating Assessment.

Several kinds of information were being sought. The four categories listed below will be reported separately in the results:

First, I wanted a comparison between university students in general and RMC students on a well known personality test (16PF).

Second, I wanted to see if at the time of recruiting the eventual leavers could be discriminated from stayers on the basis of personality, or biographical information. (16PF, PRF, and PPBQ)

Third, I wanted to examine the value of leader performance peer ratings in identifying leavers. (PEER)

Fourth, I hypothesized that all cadets would change as a function of both maturation and the military socialization process; I also wanted to see if leavers changed in a different way than stayers. (16PF, PEER)

Personality inventories were analyzed using t-test and analysis of variance procedures, while Chi-square tests of significance were used in the examination of the Personal Preference and Biographical Questionnaire responses. These were also analyzed by cluster-analytic techniques, some of which were developed at RMC, and by traditional methods of item analysis. Data analyses were carried out by members of the department. Full reports are referenced under Bain (1972, 1973), Shields (1972, 1973), and Carpenter (1972, 1973).

RESULTS

I shall report the results by study and for each instrument employed in that particular study.

Study #1. Comparison of RMC students with Cattell's norms. Table 2 displays the means and standard deviations of the RMC sample and Cattell's normative data. In comparison with other university students, RMC students are more reserved, intelligent, stable, conscientious, tough minded, trusting, imaginative, forthright, self-assured, conservative, self-sufficient, and controlled.

Study #2. Selection data relationship with stay/leave criterion.
1. 16PF. Table 3 shows the means for stayers and leavers at time of recruiting. There are three significant differences. Stayers are more self-assured, controlled, and relaxed than leavers.

2. PRF. Table 4 shows the means of stayers and leavers at time of recruiting. There are no significant differences.

3. PPBQ. Analysis of the PPBQ has provided the most interesting results to date. Using both cluster analytic and correlational techniques, three scales (voluntary, military, and academic) were produced and both methods produced multiple correlations of approximately .60 with the stay/leave criterion, and a correlation of .47 between the voluntary scale alone and the S/L criterion. The stability of this result is being tested by cross-validation and the results so far are shown in Table 5.

Study #3. Differences between stayers and leavers on PEER Ratings. Table 6 shows the mean scores on the 15 critical requirements of leadership for each of the two groups. Stayers exceed leavers on all 15, and 11 of these reach significance.

Although the PEER rating is of value in identifying potential voluntary withdrawals, it cannot be used as a screening device at the Recruiting Centre because a period of about two weeks' intimate acquaintance is required before peers have any substantial information on which to make judgements.

Study #4. Change as a function of maturation and socialization. Table 7 shows the results of three administrations of the 16PF at one year intervals. All the changes that are going to take place (with one exception) happen in the first year. Cadets become more "affected by feelings", "happy-go-lucky", "expedient", "tender-minded", "suspicious", "imaginative", "apprehensive", "undisciplined", and "tense". The one exception is "forthright", where on the third testing the group bounced back toward the "shrewd" end of the scale.

Differential changes between stayers and leavers are as follows:

1. 16PF. Table 8 provides this information in the Groups X Trials interaction. The two groups differ differentially on the forthright-shrewd factor. Leavers become more forthright while stayers become more shrewd.

2. Peer Rating. Table 9 shows that only one Leadership critical requirement is significant on the G X T interaction, ie, "interest and motivation". On it, leavers decline while stayers rise. These data were replicated using the November and April peer ratings.

DISCUSSION

In the field of testing, a significant relationship is not necessarily a useable relationship. None of the personality tests provided anything that a serious researcher would recommend to his patron as a selection device. The peer rating assessment is assuredly of value in identifying cadets who are having problems, but must be used as a counselling rather than a selection instrument because of the administration lag time.

The PPBQ alone appeared to have promise as a selection instrument. The original correlation between test and stay/leave shrank from .47 to .20 in cross-validation.

CONCLUSION

Personality factors have an elusive relationship with real-world measurable criteria. Despite the lack of useful relationships in this study we cannot abandon that strong intuitive belief that a great deal of the variance in success and failure can be accounted for by personality variables, and that given a more appropriate research design and an adequate criterion the importance of personality factors could be demonstrated. The discovery that the PPBQ, a "home-made" test was the best in the battery is a heartening experience, and a recommendation for test constructors to persevere. Further, when it is considered that candidates for RMC are a highly selected sample, relationships within that sample are sure to suffer from the effects of restriction of range.

Finally, the development of novel methodological approaches to the problem of using non-linear (categorical) data to predict categorical criteria is potentially of great value in a wide variety of prediction situations.

References

Bain, E.H. Personality, Leadership, and Motivational Characteristics of ROTP Resignees from RMC. Report 72-8, Dept of Military Leadership and Management, Royal Military College of Canada, Kingston, Ont. 1972.

Bain, E.H. The Socialization Process of ROTP Cadets at RMC - A Study of Personality, Leadership, Values, and Motivational Characteristics of Stayers and Leavers. Report 73-3, Dept of MLM, RMC, Kingston, Ont. 1973.

Carpenter, G.J. The MLM Data System. Report 72-6, Dept of MLM, RMC, Kingston, Ont. 1973.

Carpenter, G.J. A Systematic Study of the Personality, Leadership and Motivational Characteristics of ROTP Resignees from RMC. Annual Report to Defence Research Board of Activity under DRB Grant #9465-12. 1973.

Shields, W.S. A Clustering Model for Item Analysis of a Psychological Instrument with Missing Data. Educational and Psychological Measurement. In Press. (MLM Report #72-7)

Shields, W.S. The Use of Qualitative Information in the Prediction of College Attrition. Proceedings of the Annual Conference of the Canadian Association of Administrative Sciences. Queen's University. 1973.

TABLE 1

TESTS AND TIME OF ADMINISTRATION

	INSTRUMENT	INITIAL TEST	RETEST
1.	Cattell's 16PF	Recruiting Unit or 1st week at college	On release or as control subject, and annually.
2.	Jackson's PRF	Recruiting unit or 1st week at college	On release or as control subject.
3.	PFBQ	1st week at college	
4.	RMC Cadet Peer Rating	After 2 weeks at college (Aug)	Nov, Apr

TABLE 2

t-TESTS OF SIGNIFICANCE BETWEEN RMC'S AND CATTELL'S NORMS FOR THE 16PF FORM A - MALE

	Factor	RMC Sample N = 365 Average Age = 19 Years		Cattell's Sample N = 1105 Average Age = 21 Years		t
		Mean	Standard Deviation	Mean	Standard Deviation	
1.	Reserved - outgoing	8.93	3.03	9.80	3.43	-4.35***
2.	Less intelligent - more intelligent	9.63	1.59	7.72	1.80	19.1***
3.	Affected by feelings - emotionally stable	16.26	3.66	15.50	3.69	3.45***
4.	Humble - assertive	13.67	5.03	13.94	3.92	-1.04
5.	Sober - happy-go-lucky	15.68	4.43	15.73	4.40	-.19
6.	Expedient - conscientious	13.53	3.53	12.73	3.60	3.64***
7.	Shy - venturesome	13.49	5.95	13.01	5.24	1.45
8.	Tough - tender-minded	7.54	3.35	8.79	3.49	-6.25***
9.	Trusting - suspicious	7.44	3.11	9.47	3.13	-10.15***
10.	Practical - Imaginative	12.22	3.46	11.68	3.41	2.7**
11.	Forthright - shrewd	9.02	2.54	11.07	2.63	-12.06***
12.	Self-assured - apprehensive	9.17	3.54	10.65	3.90	-6.73***
13.	Conservative - experimenting	9.17	2.98	9.64	2.75	-2.76**
14.	Group-dependent - self-sufficient	10.90	3.29	9.97	3.50	4.65***
15.	Undisciplined self-conflict - controlled	13.62	3.34	10.14	3.07	17.4***
16.	Relaxed - tense	11.57	5.02	12.01	4.81	-1.47

*p < .05
**p < .01
***p < .001

TABLE 3

MEANS AND F RATIOS OF STAYERS VERSUS LEAVERS (VOLUNTARY "V")
ON 16PF (RECRUIT CAMP TESTING)

	Factor	Stayers N = 260 Means	Leavers "V" N = 55 Means	F
1.	Reserved - outgoing	9.03	9.15	0.07
2.	Less intelligent - more intelligent	9.65	9.36	1.47
3.	Affected by feelings - emotionally stable	16.33	16.18	0.08
4.	Humble - assertive	13.80	13.67	0.03
5.	Sober - happy-go-lucky	15.85	16.05	0.10
6.	Expedient - conscientious	13.66	13.09	1.14
7.	Shy - venturesome	13.98	13.16	0.87
8.	Tough - tender-minded	7.47	7.98	1.08
9.	Trusting - suspicious	7.30	8.09	2.94
10.	Practical - imaginative	12.37	11.89	0.88
11.	Forthright - shrewd	9.11	8.80	0.65
12.	Self-assured - apprehensive	8.92	9.98	4.51*
13.	Conservative - experimenting	9.24	9.15	0.04
14.	Group dependent - self-sufficient	10.61	11.40	2.69
15.	Undisciplined self-conflict - controlled	13.77	12.65	4.92*
16.	Relaxed - tense	11.20	13.27	8.01**

*$p \leq .05$
**$p \leq .01$

TABLE 4

MEANS AND F-RATIOS OF STAYERS VERSUS LEAVERS (VOLUNTARY "V") ON PRF FORM E (CFRSU TESTING)

Scale	Stayers N = 188	Leavers "V" N = 43	F df 1 and 229
1. Abasement	8.33	7.79	1.46
2. Achievement	12.68	12.23	1.09
3. Affiliation	10.09	10.65	0.96
4. Aggression	6.31	7.21	3.30
5. Autonomy	6.65	6.12	1.04
6. Change	9.87	10.07	0.19
7. Cognitive Structure	10.53	10.05	0.91
8. Defendence	4.65	4.95	0.43
9. Dominance	12.44	12.53	0.04
10. Endurance	11.79	11.40	0.86
11. Exhibition	7.07	8.05	2.21
12. Harmavoidance	5.33	5.51	0.00
13. Impulsivity	3.85	3.91	0.02
14. Nurturance	9.48	9.53	0.01
15. Order	10.23	10.84	0.78
16. Play	7.23	7.63	0.59
17. Sentience	7.59	7.42	0.09
18. Social Recognition	7.83	8.26	0.64
19. Succorance	6.31	6.16	0.07
20. Understanding	9.37	9.33	0.01
21. Infrequency	0.40	0.26	0.68
22. Desirability	12.54	11.86	3.20

*p < .05 **p < .01

TABLE 5

RESULTS OF PPBQ ANALYSIS*

1. Validation Sample 1970 + 1971 N = 229

	V	A	M
S/L	.466	.258	.372
M	.173	.165	
A	-.018		

Multiple correlation of .594 to predict attrition (S/L).

2. Cross-Validation 1972 N = 127 (using V scale with S/L criterion)

r = .20

Note: V = voluntary attrition scale
A = academic attrition scale
M = military attrition scale
S/L = stay/leave criterion

*W.S. Shields (1973)

TABLE 6

MEANS AND F-RATIOS OF STAYERS VERSUS LEAVERS (VOLUNTARY "V") ON 15 CRITICAL REQUIREMENTS (CR) OF LEADERSHIP (RECRUIT CAMP TESTING)

Critical Requirements	Stayers N = 261	Leavers "V" N = 56	F df 1 and 315
1. Performance under stress	3.32	3.19	4.61*
2. Preparation and planning	3.30	3.17	6.00*
3. Applying knowledge	3.42	3.32	4.43*
4. Taking action on his own initiative	3.38	3.20	7.59**
5. Acting honestly	3.78	3.61	8.61**
6. Admitting mistakes	3.43	3.38	0.86
7. Accepting criticism	3.41	3.36	1.19
8. Learning from experience	3.44	3.37	2.49
9. Complying with rules	3.53	3.36	7.34**
10. Assuming responsibility	3.41	3.26	4.50*
11. Clear/concise expression	3.32	3.21	4.13*
12. Effort in cooperation	3.74	3.60	6.62*
13. Appearance and habits	3.50	3.36	6.04*
14. Interest and motivation	3.61	3.28	31.35***
15. Completion of tasks	3.39	3.31	3.60
PEER TOTAL SCORE	57.40	55.20	11.49**

*$p < .05$
**$p < .01$
***$p < .001$

TABLE 7

MEANS AND F RATIOS OF SIGNIFICANCE SHOWING CHANGES IN STAYERS 16PF PROFILES OVER A 3-YEAR PERIOD - N = 62

	Factor	Means for Test Administration Years Aug 1970	Sep 1971	Sep 1972	F 1970-71	F 1971-72	F 1970-72
1.	Reserved - outgoing	9.34	9.18	8.85	0.20	1.01	1.57
2.	Less intelligent - more intelligent	9.31	9.71	9.52	2.78	0.54	0.55
3.	Affected by feelings - emotionally stable	16.58	15.40	15.02	4.87*	0.60	9.91**
4.	Humble - assertive	15.05	15.47	15.23	0.16	0.24	0.03
5.	Sober - happy-go-lucky	16.05	17.85	18.05	9.27**	0.16	11.78**
6.	Expedient - conscientious	13.66	12.61	12.87	4.87*	0.36	3.51
7.	Shy - venturesome	14.94	14.65	15.10	0.19	0.68	0.06
8.	Tough - tender-minded	7.15	8.21	8.44	4.74*	0.36	9.68**
9.	Trusting - suspicious	7.24	8.74	8.61	9.86**	0.08	14.15***
10.	Practical - imaginative	12.27	13.52	13.06	7.91**	1.02	2.38
11.	Forthright - shrewd	8.82	7.73	8.63	6.84*	6.41*	0.41
12.	Self-assured - apprehensive	8.23	9.69	9.24	7.48**	1.08	5.28*
13.	Conservative - experimenting	10.19	10.34	10.15	0.52	0.24	0.05
14.	Group dependent - self-sufficient	10.58	10.97	10.69	0.72	0.66	0.06
15.	Undisciplined self-conflict - controlled	13.85	12.69	12.45	7.20**	0.28	10.97**
16.	Relaxed - tense	11.00	12.77	12.92	11.32**	0.08	13.84***

*p < .05
**p < .01
***p < .001 df 1 and 61

TABLE 8

GROUPS (V LEAVERS VERSUS CONTROLS) BY TRIALS (PRE- AND POSTTEST) INTERACTION ON 16PF SCALE

Factor	Voluntary Leavers N = 51 Pretest Mean	Voluntary Leavers N = 51 Posttest Mean	Controls N = 43 Pretest Mean	Controls N = 43 Posttest Mean	Groups F	Trials F	Groups X Trials F
1. Reserved - outgoing	9.12	8.88	9.14	8.77	0.01	1.01	0.05
2. Less intelligent - more intelligent	9.55	10.00	9.60	9.74	0.12	2.73	0.69
3. Affected by feelings - emotionally stable	16.24	15.57	15.60	15.12	0.60	2.35	0.54
4. Humble - assertive	13.82	14.73	13.88	14.42	0.03	3.77	0.23
5. Sober - happy-go-lucky	16.24	17.37	15.51	16.35	0.89	7.16**	0.16
6. Expedient - conscientious	13.33	11.84	13.37	12.58	0.39	13.04***	1.16
7. Shy - venturesome	13.00	13.51	13.51	14.44	0.37	2.46	0.22
8. Tough - tender-minded	7.67	8.53	7.65	8.67	0.01	9.21**	0.09
9. Trusting - suspicious	8.16	8.25	7.56	7.70	0.95	0.12	0.00
10. Practical - imaginative	11.71	12.29	11.48	12.14	0.10	2.84	0.01
11. Forthright - shrewd	8.78	7.90	8.67	9.02	1.29	1.26	4.67*
12. Self-assured - apprehensive	9.98	10.92	9.09	9.72	3.32	5.14*	0.20
13. Conservative - experimenting	9.45	10.33	10.00	9.70	0.01	0.94	2.82
14. Group dependent - self-sufficient	11.37	11.39	10.44	11.05	1.10	0.60	0.68
15. Undisciplined self-conflict - controlled	12.94	11.63	13.65	12.63	2.39	8.30	0.13
16. Relaxed - tense	13.12	13.29	11.60	12.42	1.74	1.24	0.57

*p $\leq$ .05
**p $\leq$ .01
***p $\leq$ .001 df 1 and 92

TABLE 9

GROUPS ("V" LEAVER VERSUS STAYERS) BY TRIALS (PRE AND POSTTEST) INTERACTION ON 15 CR'S OF LEADERSHIP

Critical Requirements	Leavers N = 46		Stayers N = 264		Groups	Trials	Groups X Trials
	Aug test mean	Nov test mean	Aug test mean	Nov test mean	F	F	F
1. Performance under stress	3.21	3.20	3.29	3.37	3.52	6.56*	1.38
2. Preparation and planning	3.18	3.19	3.26	3.41	6.85**	23.47***	3.43
3. Applying knowledge	3.34	3.34	3.38	3.48	3.38	10.73**	1.59
4. Taking action on his own initiative	3.21	3.30	3.34	3.39	2.54	4.04*	0.27
5. Acting honestly	3.69	3.68	3.74	3.76	1.60	0.66	0.01
6. Admitting mistakes	3.41	3.40	3.39	3.36	0.22	0.69	0.08
7. Accepting criticism	3.38	3.36	3.38	3.35	0.02	1.01	0.00
8. Learning from experience	3.39	3.35	3.40	3.46	1.46	4.01*	2.07
9. Complying with rules	3.39	3.41	3.49	3.57	3.51	8.15**	0.78
10. Assuming responsibility	3.26	3.22	3.37	3.38	3.50	0.11	0.63
11. Clear/concise expression	3.18	3.30	3.28	3.38	2.70	15.59***	0.13
12. Effort in cooperation	3.62	3.59	3.70	3.70	2.04	0.06	0.13
13. Appearance and habits	3.36	3.37	3.46	3.58	5.23*	14.45***	1.91
14. Interest and motivation	3.33	2.99	3.57	3.60	43.25***	0.79	21.05***
15. Completion of tasks	3.32	3.30	3.35	3.47	4.51*	14.20***	3.59
PEER TOTAL SCORE	55.43	55.05	56.75	57.52	6.13*	2.59	1.18

*p $<$ .05 df 1 and 308
**p $<$ .01
***p $<$ .001

MOTIVATION OF ENLISTED RESERVISTS IN A NON-DRAFT ATMOSPHERE

BY

LT KNIGHT D. CHENEY[1]
CAREER PATTERNS AND TRAINING PROGRAMS BRANCH
RESERVE TRAINING DIVISION
U. S. COAST GUARD HEADQUARTERS
WASHINGTON, D. C.

As people involved in the training and evaluation of military personnel, we are today working in a hostile atmosphere - a time when our services face a manpower crisis both in quality and numbers. Hallowed traditions are falling by the wayside at military academies, enlisted people are asking "why?" more often and prospective military personnel remember Vietnam with some distaste. The draft has disappeared as an inducement for satisfactory military performance.

As the number of personnel on active duty dwindles, emphasis on the importance of a strong reserve force grows. Examples reflect situations in the writer's service. However, it is hoped that the concepts illustrated have more general applicability.

Definitions and Assumptions:

1. A reservist who is well trained in his rating (field of specialization) is well trained for mobilization.

2. Augmentation consists of any reserve activity that supports effective training for mobilization while meeting a stated need of an element of the regular Coast Guard.

3. Goal:

a. Provide (officers and) petty officers qualified to perform "X" effectively.

b. Retain qualified reservists.

4. "X" is defined as a composite of technical, practical and common sense traits PLUS leadership, loyalty, and sense of responsibility required to perform satisfactorily in a billet that meets a stated need of the regular component.

A great deal of work has been done in identifying the technical and practical traits mentioned in the definition of "X". Leadership schools, patriotic publications, coercion and other ways have been widely used to instill other desirable qualities. Let us assume for now that the required technical and practical skills have been identified. As trainers and evaluators we desire to (1) provide (ECONOMICALLY) enlisted reservists with the necessary learning and (2) we wish to maintain skill levels.

We desire to accomplish our training task without resorting to negative incentives such as threatening call to extended active duty as a punitive measure.

[1]The opinions or assertions contained herein are the private ones of the writer and are not to be construed as official or reflecting the views of the Commandant or the Coast Guard at large.

In the Coast Guard Reserve, we have approximately 9500 enlisted reservists in 37 ratings. They are well educated. Over 20% of the group have sixteen or more years of education. Over one-half have more than twelve years of education and over 90% have high school diplomas. These people enlisted for various reasons. Until very recently the draft has been an important motivating factor in both enlistment and continued satisfactory participation. This point is very significant for reasons which will become apparent in a few moments.

Enlisted advancements are earned through completing a variety of requirements including time-in-grade, correspondence courses and successful competition in servicewide examinations. These exams are the same as the ones given Regular personnel and are an important part of maintaining a "one service" standard.

Correspondence course completion is a prerequesite to servicewide competition.

Emphasis is placed on servicewide examinations here because it gives us a relatively objective measure of training needs by topic area. Further, it is felt that level of participation is a rough indicator of interest in the program.

Participation in correspondence courses and servicewide examinations has fallen sharply in the last two years. Reasons for the decline include the near demise of the Coast Guard Reserve in 1971, which caused morale problems, and the trend away from traditional monthly drills in a classroom atmosphere - the advent of augmentation.

Enlisted personnel in the Coast Guard Reserve are distributed through the pay grades as follows:

TABLE I

Distribution of Enlisted Personnel in the Coast Guard Reserve (Selected), Percentage by Pay Grade, 30 July 1973.

Rate	Percentage	Desired Percentage
E-4 and Below	53.32	40
E-5	21.87	30
E-6	14.97	20
E-7 and Above	9.84	10
	100.00	100

This summer, 1700 questionnaires were distributed to enlisted reservists. 538 were returned and the proportion returned were roughly equivalent by pay grade to on-board strength. Table Two indicates response to pertinent questions by pay grade.

TABLE II

Note several interesting things:

1. The higher the pay grade, the greater the faith in the advancement system (reflected in the question "Do you think the servicewide exam system is fair?").

2. The relatively low participation in servicewide exams in April 1973.

3. Indicated applicability of using civilian vs. military related skills.

4. Reenlistment intentions by pay grade.

Now, recall the personnel distribution by pay grade and compare with the reenlistment intentions represented here. Note that the LOWEST expected retention rate is in the most highly populated pay grades which, in turn serve as a source for pay grades E-5 and E-6 where we have comparative deficits.

Enlistment dates indicate that many of the personnel in the E-4 and below category enlisted while the draft was still active.

The obvious challenge is retention in the lower pay grades and the encouragement of advancements. The special target is the group of reservists who anwered "Maybe" to the question "Doy you plan to reenlist?".

Now, let us look at a pyramid presentation of Maslow's hierarchy of needs. Proposal 1. It is proposed that the typical reservist's position in the hierarchy has shifted upward (or, in the case of obligors, will shift upward upon completion of required service) in the absence of the draft and that this factor must be addressed if a satisfactory climate for personal performance and retention is going to exist.

TABLE III

Dilemma: This approach must involve a commitment on the part of the program's administrators as well as a change in written directives, guidelines and other communications. Proposal 2. Service training requirements can be met in a way that also contributes to the reservist's personal needs. The concept of working toward BOTH goals should support quality participation and retention.

Recall the definition of augmentation. - "Augmentation consists of any reserve activity that supports effective training for mobilization while meeting a stated need of the regular Coast Guard." A recent survey of enlisted reservists in the 11th Coast Guard District measured reenlistment prospects for obligated and non-obligated reservists by type of augmentation duty. The varieties of augmentation that equated with highest retention potential among obligors, especially, were those that involved active identifiable accomplishment - boating safety and search and rescue.

TABLE IV

Less substantial activities did not enjoy the same implied enthusiasm. A fringe benefit of augmentation is that productive work is being accomplished - not typically the case in the past when reservists spent drill after drill

in classrooms sleeping through WW II training films.

Admittedly, augmentation alone is not likely to produce the variety of skills required to attain "X" as defined earlier. Other types of training must be included in the blend to attain the desired result. This mix must contribute to the reservist's needs as well as military requirements if we are to attain our goals.

Let us now look at some specific things that can be done to accomplish the ends set forth:

1. Install, use and maintain an effective automated personnel accounting system. This management tool offers:

a. Ready availability of data to personnel action staffs.

b. A convenient index of available reservist skills.

c. Elimination of redundant personnel procedures such as repeated requests for the same information.

d. Ready identification of refresher training needs.

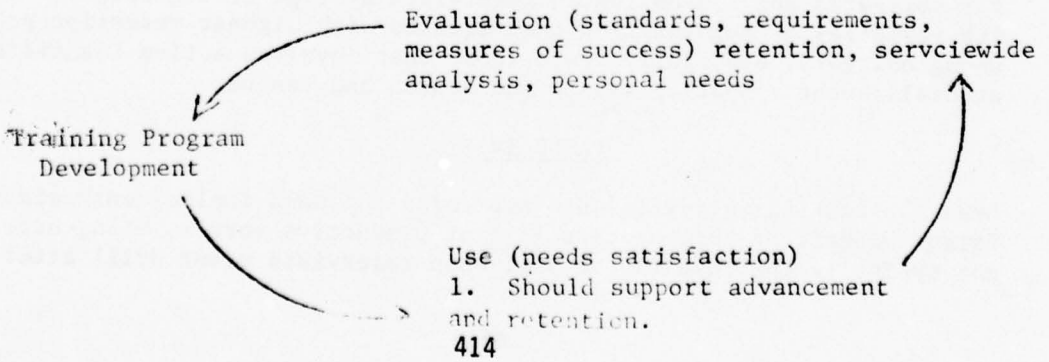
e. Predictor information for advancement eligibility, retirements, etc.

2. Take advantage of getting reservists off to a positive start by insuring quality programs and well trained instructor personnel at initial training centers.

3. Personalize the reservist's relationship with his service through expanded use of direct mailing, prompt attention to correspondence and more direct contact with officer/advisors. Does the reservist have civilian skills compatible with service needs? Let's use them unless the person involved really wants variety. (Assumption - adequate information is available to the reservist for him to make an intelligent decision). We can also provide for changes in field specialization where justified. Here is a real opportunity to employ a great, relatively untapped resource, the reserve junior officer.

4. Periodically review all training requirements to insure that they are relevant and realistic. One way this may be accomplished is by taking the following approach:

TABLE V



For example, servicewide exam analysis by section indicated a need to strengthen skills in "military factors" or knowledge in areas applicable to all enlisted personnel in all ratings. Military factors cover such topics as UCMJ, first aid, customs and courtesies, etc. and are traditionally regarded as "Mickey Mouse". Knowledge of military factors enhances a reservists change for earning a qualifying score on the servicewide exam. A way to learn the required items is being made available through use of a highly portable audio-visual training device together with a quiz.

5. Maximize use of "where the action is" training resources and improve effectiveness of other resources.

6. Offer, so far as is possible, an equitable opportunity for advancement and opportunity to earn retirement benefits.

Summary:

Enlisted reservists are an intelligent group of individuals whose skills are becoming more important as regular force resources become more restrained.

A desirable goal is to provide enlisted reservists with technical, practical and other skills that will allow them to perform satisfactorily in specified billets. In addition, training should contribute to retention by meeting reservists personal needs on a higher plane than in the past.

Tools such as effective automated personnel accounting systems, getting reservists off to a good start in their service careers, personalizing relationships with the service, guaranteeing, when possible, equitable advancement opportunities, a chance to earn retirement benefits and evaluating training programs with an eye toward whether they contribute to human needs, will help keep our reserve forces strong and viable.

BIBLIOGRAPHY

Abraham H. Maslow: Motivation and Personality, 2nd edition, Harper and Row, New York, Pp. 35-58, 98-104

T. O. Jacobs: Leadership and Exchange in Formal Organizations, Human Resources Research Organization, Alexandria, Virginia 1971, Pp. 101, 122-154.

Frederick J. McDonald and Garlie A. Forehand: A Design for Accountability in Education, New York University Education Quarterly, Volume IV No. 2, Winter 1973, Pp 7-15.

Walter S. Lee: The Assessment, Analysis and Monitoring of Educational Needs, Educational Technology, Volume XIII, No. 4, April 1973, Pp 28-32

TABLE I

Distribution of Enlisted Personnel in the Coast Guard Selected Reserve, Percentage by Pay Grade, 30 July 1973.

RATE	PERCENTAGE	DESIRED PERCENTAGE
E-4 and Below	53.32	40
E-5	21.87	30
E-6	14.97	20
E-7 and Above	9.84	10
	100.00	100

Source: CCGD11(r) ltr 7430/mmd dtd 13 Sept 1973

TABLE II

DO YOU THINK THE SERVICEWIDE EXAM SYSTEM IS FAIR?

Pay Grade	E-3	E-4	E-5	E-6	E-7	Combined
YES	37.04%	45.44%	55.32%	58.70%	77.36%	53.16%
NO	18.52%	25.13%	21.99%	14.13%	13.21%	20.26%
NO RESPONSE	44.44%	29.32%	22.70%	27.17%	9.43%	26.58%

DID YOU COMPETE IN THE APRIL 1973 SERVICEWIDE EXAMINATION?

Pay Grade	E-3	E-4	E-5	E-6	E-7	Combined
YES	5.56%	7.73%	10.64%	11.96%	18.87%	10.04%
NO	94.44%	92.15%	84.40%	88.04%	81.13%	89.96%

DO YOU FEEL THAT YOU HAVE A REASONABLE CHANCE OF COMPLETING SUFFICIENT TIME IN THE COAST GUARD RESERVE FOR RETIREMENT?

Pay Grade	E-3	E-4	E-5	E-6	E-7	Combined
YES	53.70%	51.84%	75.18%	95.65%	98.11%	71.00%
NO	46.30%	44.50%	19.86%	3.26%	1.89%	26.39%
NO RESPONSE		3.66%	4.96%	1.09%		2.61%

DO YOU FEEL THAT YOU COULD CONTRIBUTE MORE TO THE COAST GUARD BY USING CIVILIAN JOB RELATED SKILLS TO A GREATER EXTENT?

Pay Grade	E-3	E-4	E-5	E-6	E-7	Combined
YES	62.96%	56.55%	66.67%	65.22%	73.58%	63.20%
NO	37.04%	39.79%	31.91%	30.44%	26.42%	34.57%
NO RESPONSE		3.66%	1.42%	4.35%		2.23%

IF RESPONSE TO THE ABOVE QUESTION WAS YES, WOULD YOU

Pay Grade	E-3	E-4	E-5	E-6	E-7	Combined
PREFER TO USE CIVILIAN JOB RELATED SKILS	48.15%	39.27%	35.46%	65.22%	73.58%	37.17%
PREFER THE VARIETY I GET BY DOING SOMETHING DIFFERENT AS A RESERVIST	20.37%	30.37%	41.13%	30.44%	26.42%	36.25%
NO RESPONSE	31.48%	30.37%	23.41%	4.35%		26.58%

DO YOU PLAN TO REENLIST?

Pay Grade	E-3	E-4	E-5	E-6	E-7	Combined
YES	11.11%	19.90%	35.46%	71.74%	98.11%	40.71%
NO	35.19%	43.46%	21.28%	2.17%	1.89%	24.72%
MAYBE	53.70%	36.65%	43.26%	26.09%		34.20%

SOURCE: USCG HQ (G-RT) RECORDS OF RESULTS OF SUMMER '73 RESERVIST QUESTIONNAIRE

TABLE III

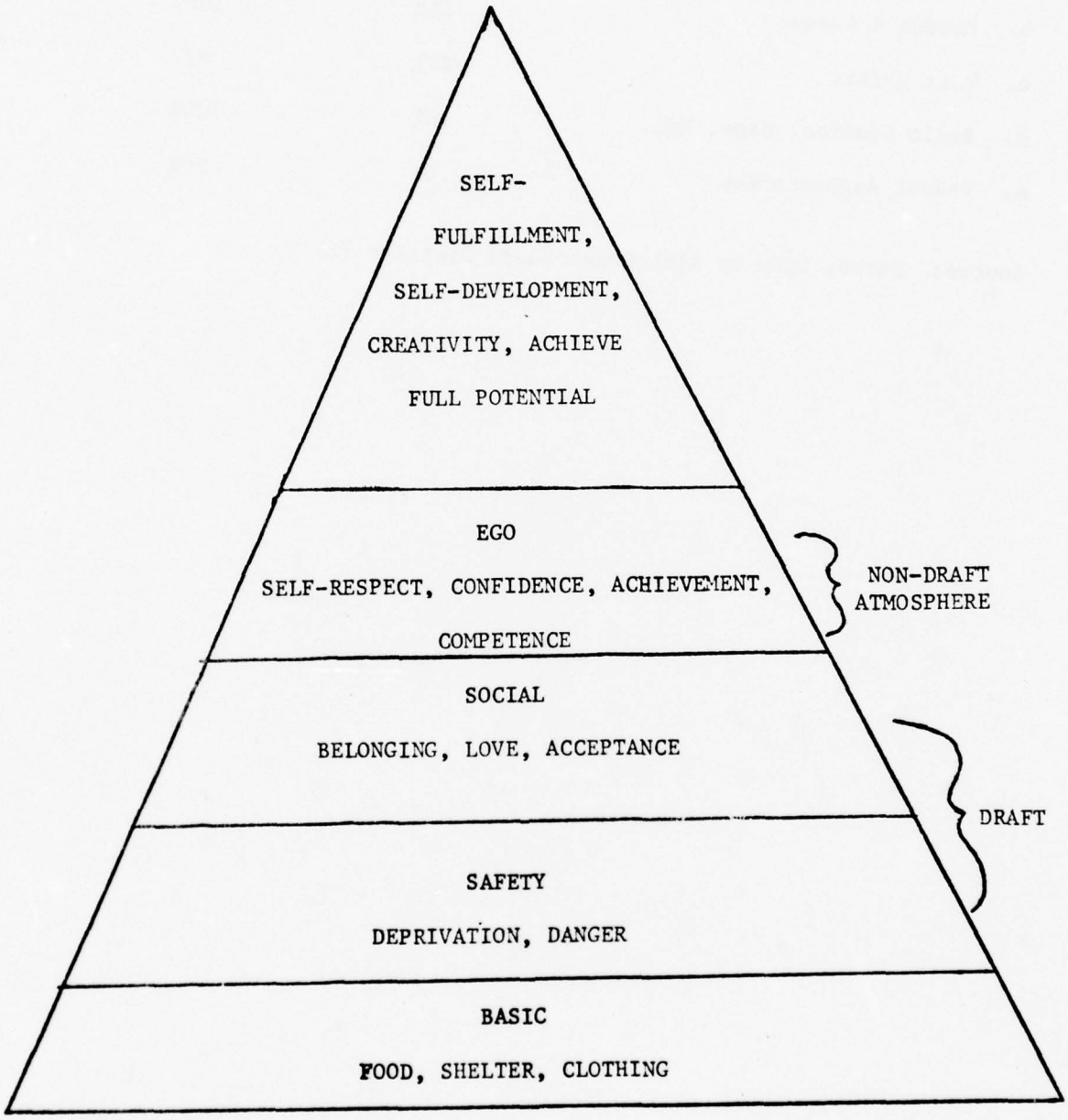

SELF-
FULFILLMENT,
SELF-DEVELOPMENT,
CREATIVITY, ACHIEVE
FULL POTENTIAL
EGO
SELF-RESPECT, CONFIDENCE, ACHIEVEMENT,
COMPETENCE
NON-DRAFT
ATMOSPHERE
SOCIAL
BELONGING, LOVE, ACCEPTANCE
DRAFT
SAFETY
DEPRIVATION, DANGER
BASIC
FOOD, SHELTER, CLOTHING

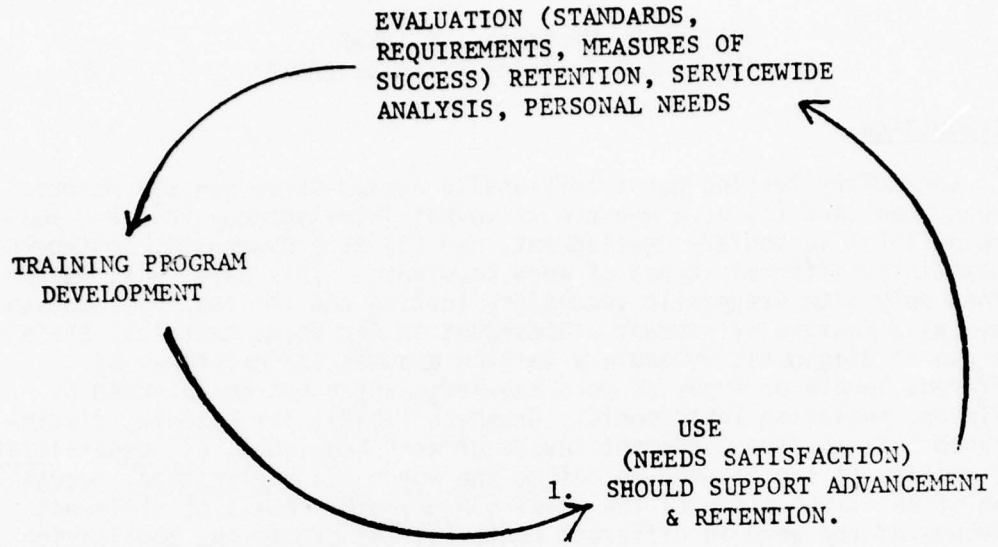
TABLE IV

REENLISTMENT POTENTIAL BY MISSION

MISSION	OBLIGORS	NON-OBLIGORS
a. Boating Safety	65%	92%
b. Search & Rescue	41%	100%
c. Port Safety	27%	97%
d. Radio Station, Base, Etc.	25%	100%
e. Vessel Augmentation	6%	90%

Source: Survey made by 11th Coast Guard District (r)

TABLE V


EVALUATION (STANDARDS, REQUIREMENTS, MEASURES OF SUCCESS) RETENTION, SERVICEWIDE ANALYSIS, PERSONAL NEEDS
TRAINING PROGRAM DEVELOPMENT
USE
(NEEDS SATISFACTION)
1. SHOULD SUPPORT ADVANCEMENT & RETENTION.

An Evaluation of Diagnostic Vocabulary Testing in Air Force Technical Training

William P. Mockovak, Lt, USAF
AFHRL Technical Training Division (AFSC)

Introduction

Vocabulary testing has traditionally served three general purposes; it has been used (1) as a measure of verbal intelligence, (2) as a measure of total vocabulary development, and (3) as a diagnostic instrument in examining different types of word knowledge. This paper will be concerned only with diagnostic vocabulary testing and its possible applications as a measure of student achievement in Air Force technical training. The use of diagnostic vocabulary testing assumes the existence of different levels or types of word knowledge which can be detected by different measuring instruments. Cronbach (1942), for example, distinguished between five different levels of word knowledge, (1) generalization, the ability of the student to define the word; (2) application, recognition of an illustration of the word; (3) breadth, recall of different meanings of the word in different contexts; (4) precision, application of the word to all possible situations - even unfamiliar ones; and (5) availability, the student's use of the word in thinking and discourse. Generally, diagnostic vocabulary testing has been concerned with the breadth and precision of word knowledge and the test forms most commonly used have been recall, multiple choice, matching for correct answer, synonym, or example, and the checklist.

If diagnostic vocabulary testing is to be employed as a measure of student proficiency in technical training, there are two assumptions which must first be satisfied. The most important of these is that technical vocabulary development must be at least positively related to other measures of technical course proficiency or achievement, although the magnitude of the relationship need not be great. Intuitively, this would seem to be the case, since in technical training, especially, knowledge of technical words appears to be an essential prerequisite for communication and skill acquisition whether it be through a textbook or an instructor. Another assumption is that some valid measuring instrument can be developed which accurately reflects the level of an individual's technical vocabulary development, either in an absolute or diagnostic sence. Theoretically, there appears to be additional justification for technical vocabulary testing. Langer (1967), in discussing the relationship between vocabulary and concept development, states that, "the development of vocabulary is an inextricable part of concept development." Within this context, words form the basis for the classification of concepts since they are the "labels" which are cognitively manipulated by individuals in thinking and reading. If, in fact, the preceding statement is true and some studies suggest that it is, then there is ample justification for testing word knowledge as a means of measuring technical proficiency.

For example, the results of a study conducted by Randleman (1963) indicated that the size of technical vocabularies was positively related to measures of achievement such as IQ and scholastic ability; however, his results also indicated that vocabulary growth is attributable to factors other than intellectual ability.

The development of valid test instruments to measure vocabulary development is another problem in itself. Multiple choice tests have been used most frequently because of their ease in administration and scoring but there are distinct drawbacks with this approach. For one thing, studies have shown that risk-taking behavior may affect test scores (Slaleter and Koehler, 1968; Ziller, 1957). Also, different scoring instructions have been found to influence a student's overall performance. For example, Waters and Waters (1971) used a multiple choice format in an effort to influence risk-taking behavior. Varying the test instructions resulted in the following scoring procedures: (1) a fractional amount was subtracted for a wrong answer, (2) a fractional amount was added for omitted items, and (3) a condition where there was no reward and penalty. They concluded that if guessing is to be discouraged on a multiple choice test, a procedure for rewarding the examinee for omitted items would be superior to penalizing the examinee for wrong answers. In addition to scoring instructions, the choice of which particular words are to be included in multiple choice vocabulary tests is also an important consideration since abstract words, such as adjectives, adverbs, and verbs have traditionally been employed in vocabulary tests as a means of estimating verbal intelligence (Cureton, 1963). On the other hand, a technical vocabulary test is interested in measuring achievement, not verbal intelligence; therefore, appropriate vocabulary selection must be an important consideration.

With the preceding considerations in mind, the use of multiple choice testing to assess technical vocabulary development must be approached with caution. However, possibly the most severe criticism of multiple choice testing when attempting to discriminate between different levels of technical knowledge is that under standard scoring procedures, an answer is either right or wrong, despite varying degrees of student knowledge. For example, if four students were asked to select the proper alternative for the word "polling" (Question 21, Attachment A), and students A and B got the correct answer, but students C and D did not, a supposedly valid assumption would be that students A and B had equal states of knowledge. However, it is quite possible that student A had no knowledge of the question or alternatives and merely guessed correctly; whereas, student B had sufficient knowledge of the term and alternatives to pick the correct answer. It is also feasible that student C had partial knowledge of the term and was able to narrow his choice down to two alternatives, at which point he guessed incorrectly. Student D, on the other hand, may have been

convinced that the answer he picked was right and all the others were wrong. Obviously, the four students have varying states of knowledge which a multiple choice test is not sensitive enough to detect. Questions must be raised therefore about the validity of the multiple choice test as a sensitive discriminator of student knowledge.

In an attempt to overcome the shortcomings of traditional multiple choice testing, confidence testing has been proposed as a means of determining a student's degree of confidence in various testing situations. Simply defined, "Confidence testing is a method of testing which provides for weighting, either directly, or indirectly, a selected alternative or alternatives in such a way as to reflect the examinee's belief in the correctness of his response" (Echternacht, 1971). It is hoped, of course, that the score obtained using confidence testing would validly reflect varying degrees of knowledge since an important component - the student's confidence in his answers - is being included. However, there are two important assumptions which must be satisfied before confidence testing should be used; (1) it must be assumed that the examinee is interested in obtaining a high score, and (2) the scoring rule must be known to the examinee. There have been many claims made on the behalf of confidence testing such as it increases reliability, it relieves student anxiety in testing situations, it is a "fairer" test, etc., but as Echternacht has stated, the ultimate value of confidence testing can only be judged by the gain in "ability" variation against the gain in "error" variation. The increase in error variation can be attributed to such things as (1) risk-taking behavior patterns in expressing confidence, (2) confusion about the scoring procedure, (3) inability to express accurate estimates of subjective probabilities of knowledge, and (4) a lack of desire to perform well. Echternacht and Boldt (1971) have reviewed four confidence testing procedures which have received a significant amount of attention. They are, (1) Pick One, (2) Distribute 100 Points, (3) Sco-Rule, and (4) Five Stars. These procedures differ primarily in the assignment of weights to either one or all of the alternatives; however, there are also significant differences in student/instructor preferences and administration and scoring time. It is suggested that the interested reader refer to the review for a more thorough discussion of the relative strengths and weaknesses of the different confidence scoring procedures.

In light of the previous studies, there appears to be theoretical and empirical justification for using technical vocabulary measures as indicators of student achievement in technical training. Additional empirical support is required but the most pressing consideration involves the development and validation of appropriate test instruments. A multiple choice test form with a confidence scoring procedure offers the greatest potential because theoretically it will result in finer discriminations of student knowledges which should have a significant

impact on remediation and the conduct of instruction. However, it is also possible that the traditional multiple choice test form with "rights only" scoring will accomplish the same purposes with less administrative effort. In an effort to clarify this issue, as well as related questions, this study attempted to accomplish the following objectives; (1) to determine the "feasibility" of using technical vocabulary testing in technical training, (2) to determine if Diagnostic Vocabulary Testing is a valid means of assessing technical course proficiency, (3) to compare confidence testing and multiple choice scoring formats in order to determine if the procedures result in differential amounts of information concerning student proficiency, and (4) to determine the extent of the relationship between reading ability, intelligence, and acquisition of technical vocabularies.

Procedure

In the present study, a technical vocabulary test was developed which employed the Pick One confidence testing procedure. This method was chosen because it is relatively easy to explain to students, it is easy to score, and testing time is only slightly longer than traditional multiple choice (Echternacht et al, 1971). The scoring scheme was chosen in such a manner that wrong opinions confidently expressed incurred large penalties, frank guesses and near misses were only mildly punished or rewarded if at all, and confidently expressed correct opinions were greatly rewarded. However, the scoring scheme was also developed so that a student would score highest if he gave an honest and accurate estimate of his degree of certainty. This characteristic of the scoring scheme is termed "reproducibility" and it is discussed in greater detail by Boldt (1971).

The technical vocabulary test (TVT) (Attachment A) was developed for the Inventory Management Course (3ABR64530-1) which is taught on a resident basis at Lowry AFB. This particular course was selected because it contained a sufficient amount of technical jargon, yet the vocabulary was not so difficult that someone who was not an expert in the area could not develop a pool of vocabulary items. Technical vocabulary was extracted from all sections of the course to develop a representative pool of vocabulary items and from this pool 35 items were selected to represent the major parts of the course. Each item had four alternatives, an example of the answer sheet is presented in Appendix B.

The TVT was reviewed and corrected for technical accuracy by course instructors and administered to 40 students who had just graduated from the 4-block course of study. Before administering the test, the confidence scoring procedure was explained to the

students and several sample problems were worked through with the entire class. The weighting system used follows in Table 1 (from Echternacht and Boldt, 1971):

Table 1

	Degree of Confidence				
	Very Sure	Sure	Fairly Sure	Not Very Sure	Not Sure
Credit if right	10	9	7	3	0
Penalty if wrong	17	12	5	2	0

NOTE: The instructions stressed the fact that a student would score his highest if he honestly expressed his confidence.

In addition to the TVT, the students were also administered the United States Armed Forces Institute Achievement Test III (UAT III), Part I (Vocabulary Sub-Test). The students' scores on this test were converted to estimated reading grade levels (RGLs) using a regression equation, since there is a high correlation (.82) between reading scores and vocabulary scores on the UAT III.

Results

AQE and AFQT data were available for 30 of the 40 students who took both vocabulary tests. The average scores were as follows (standard deviations in parenthesis): Gen AQE = 58.7 (11.0), Mech AQE = 49.5 (18.7), Admin AQE = 73.3 (12.1), Elec AQE = 59.7 (13.4), and AFQT = 56.0 (17.5). The scores on the TVT were obtained using the confidence weighting scheme and also the traditional right-or-wrong (rights only) scoring procedure. It was assumed that the answer a student selected with the confidence testing instructions would have been the same answer selected with standard multiple choice instructions. The correlations of scores on the TVT (with both methods of scoring) with final course grades were significant but the confidence scoring procedure did not yield a significantly higher correlation than standard multiple choice testing. In fact, confidence testing resulted in a slightly lower correlation with final course grades than multiple choice, .42 vs .44 (Table 1). (A student's final course grade was calculated by averaging his end-of-block test scores for the 4-block course.)

Table 2

Intercorrelations of Student Measures

	TVT (Mult Choice)	TVT (Conf Testing)	Final Course Grade
UAT III (RGL)	.46*	.53**	.26
TVT (Mult Choice)		.94**	.44*
TVT (Conf Testing)			.42*

N = 40
* $p < .05$
** $p < .001$

Nevertheless, the significant correlations obtained indicate that the TVT is tapping a portion of course knowledge.

Another question which the data in Table 1 partially answer concerns the amount of additional predictive information obtained using the confidence scoring procedure. The two scoring procedures used on the TVT correlated .94 with each other, which indicates that from a psychometric point of view, the additional effort involved with confidence testing may be unnecessary because there is no practical difference in the scoring systems.

Referring to Table 1, once again, it is also interesting to note the significant correlation between reading grade level (RGL) and scores on the TVT using either scoring procedure. These data would seem to indicate that reading ability plays a part in the acquisition of a technical vocabulary; however, when the standardized reading scores were partialed out of the TVT (Mult Choice) scores and final course grades, the TVT still correlates. 38 ($p < .05$) with final course grades which indicates that the TVT is not simply a measure of reading ability.

Another question which deserves attention involves the degree of relationship between the TVT and measures of aptitude (AQE) and general intellectual ability (AFQT). Table 2 presents the correlations between individual AQE and AFQT scores and scores on the TVT. A smaller sample size (N = 30) was available in this case because of missing student data.

Table 3

	TVT (Mult Choice)	TVT (Conf Testing)
Mech AQE	.26	.25
Gen AQE	.27	.39*
Admin AQE	.22	.28
Elec AQE	.14	.18
AFQT	-.05	.05

N = 30
* $p < .05$

In addition, the lower sample size reduced the score variance (TVT) which probably resulted in lower correlations. The only significant correlation obtained was between Gen AQE and TVT (Conf Testing). The non-significant correlation between AFQT and TVT is especially interesting because it may indicate that technical vocabulary development is independent of general intellectual ability as measured by such tests as the AFQT. This possibility was suggested by Langer (1967) in that he views concept formation ability as being independent of intelligence. Actually, however, scores on technical vocabulary tests in general, and the TVT used in this study, specifically, should not correlate highly with "pure" intelligence measures since most technical vocabulary consists of "concrete" words, rather than "abstract" words which have been found to correlate highly with verbal intelligence (Cureton, 1963).

Discussion

In general, the results of this study seem to indicate that technical vocabulary testing is a feasible approach towards measuring a particular aspect of student achievement in Air Force technical training. A technical vocabulary test is easy to develop and it requires relatively little administration time; however, the question arises of how it can be used in conjunction with standard student achievement measures such as course quizzes, end-of-block exams, etc. It would not be desirable to use a TVT alone as an end-of-block exam because its scope is not broad enough to cover the diverse skills required in most courses; however, it could be used to supplement block exams as an additional measure of student achievement. Because of its short administration time (20-30 minutes for a 35-item test), however, a TVT could be used quite easily as a remedial aid, particularly if the vocabulary items are chosen with care to reflect essential course knowledges. An instructor could refer quite readily to such a test in order to determine where confusions in essential

concepts might be occurring. Unfortunately, the confidence testing procedure used in this study did not significantly increase the amount of information available from the TVT. Such information would have been useful in diagnosing student confusions, but on the basis of other studies (Echternacht et al, 1971), it appears that confidence testing is not very feasible as a diagnostic aid in technical training. In the preceding study, confidence testing was employed with regular course quizzes and different types of remediation in two technical courses in an effort to determine if an implementation of confidence testing procedures affected end-of-block grades, number of remedial sessions, student and instructor attitudes, and administration and scoring time. The experimental design confounded the results of implementing confidence testing but from instructor's responses and attitudes, it was clear that they tended not to use confidence testing in planning remediation. As one instructor remarked, "It reminds me of using a bulldozer to clear snow from a sidewalk -- it's too good." Also, in this study, Pick One confidence testing correlated .92 and .89 (compared to .94 in the present study) with "rights only" multiple choice scoring, so once again there appears to be little practical difference in the scoring systems other than changing the scale of measurement. One feature of the Echternacht et al study which does deserve more attention, however, is the use of special types of remediation in conjunction with quizzes or TVTs, since confidence testing combined with special remediation did reduce the number of student remedial sessions.

Another purpose of the present study which deserves discussion involves the relationships between reading, aptitude, and intelligence measures with scores on the TVT. The fact that reading ability correlated significantly with the TVT was expected, but the low and non-significant correlations of AFQT (Table 3) with the TVT indicate that technical vocabulary acquisition may be a specific aptitude in addition to the other aptitude measures employed. It is true that the Gen AQE correlated significantly with the TVT (Conf Testing); however, when the standardized reading scores (UAT III) are partialed out of Gen AQE and the TVT, the correlation between Gen AQE and TVT (Conf Testing) drops to a non-significant .24. Again, this suggests that technical vocabulary acquisition is a factor relatively independent of general intellectual abilities. If this hypothesis can be substantiated through further experimentation then the possibility exists that training innovations can be introduced which will help individuals acquire technical vocabulary and perform more effectively in their career specialties.

References

1. Boldt, Robert F. A Simple Confidence Testing Format. AFHRL-TR-71-31. Lowry AFB, Colorado: Technical Training Division, Air Force Human Resources Laboratory, July 1971

2. Cronbach, Lee J. An Analysis of Techniques For Diagnostic Vocabulary Testing. Journal of Educational Research, 1942, 36, 206-217

3. Cureton, E. E. Note On Vocabulary Test Construction. Educational and Psychological Measurement, 1963, 23, 461-466

4. Echternacht, G. J. Use of Confidence Testing In Objective Tests. AFHRL-TR-71-32. Lowry AFB, Colorado: Technical Training Division, Air Force Human Resources Laboratory, July 1971

5. Echternacht, G. J. and Boldt, R. F. User's Handbook for Confidence Testing As A Diagnostic Aid In Technical Training. AFHRL-TR-71-34. Lowry AFB, Colorado: Technical Training Division, Air Force Human Resources Laboratory, July 1971

6. Echternacht, G. J.; Sellman, W. S.; Boldt, R. F.; and Young, J. D. An Evaluation of the Feasibility of Confidence Testing As A Diagnostic Aid In Technical Training. AFHRL-TR-71-33. Lowry AFB, Colorado: Technical Training Division, Air Force Human Resources Laboratory, July 1971

7. Langer, J. H. Vocabulary and Concept Development. Journal of Reading, 1967, 10, 448-456

8. Randleman, R. R. Technical Vocabulary and Its Relationship to Selected Student Characteristics And Industrial Education Programs. American Vocational Journal, 1963, 38, 27

9. Slaleter, M. J. and Koehler, R. A. New Measure of Risk Taking on Objective Examinations. California Journal of Educational Research, 1968, 19, 132-137

10. Waters, C. W. and Waters, L. K. Validity and Likability Ratings For 3 Scoring Instructions For A Multiple-Choice Vocabulary Test. Educational and Psychological Measurement, Winter 1971, 31, 935-938

11. Ziller, R. C. A Measure of the Gambling Response-Set in Objective Tests. Psychometrika, 1957, 22, 289-292

INVENTORY MANAGEMENT TECHNICAL VOCABULARY ASSESSMENT

1. Issue

   (a) An actual equipment item obtained from supply

   (b) A sale under the stock fund concept

   (c) The procedure for obtaining an equipment item from supply

   (d) A computer process

2. Grave Damage

   (a) Results when destruction of equipment is over $5,000.

   (b) Results when unauthorized disclosure of Top Secret information.

   (c) Results when an aircraft is grounded due to lack of parts.

   (d) Results when a base is closed.

3. Master Item

   (a) An item which must be kept in a controlled area

   (b) An item which best meets the needs of the AF

   (c) An item composed of many sub-components

   (d) An item used only rarely

4. AFS

   (a) Air Force Supply Career Field

   (b) Code for Army Federal Stock Number

   (c) Positions that require common qualification with a career field

   (d) None of the above

5. SNUD

   (a) Indicates the age of an item

   (b) A central file at Hq ATC

   (c) A central file at Hq USAF

   (d) A central file at Hq AFLC

6. Bench stock

   (a) Common items used to repair end items

   (b) rare items used to repair end items

   (c) Large storable items

   (d) Extremely large equipment pieces

7. KIT

   (a) Parts used to accomplish a complete modification on an item of equipment

   (b) A communication asking for information from the computer

   (c) A package of related parts

   (d) Refers to a delayed transaction

8. Pre-inventory

   (a) Preparations made for conducting physical counts of items

   (b) A request for items

   (c) A physical count of a certain item prior to counting all items

   (d) The name of a computer component

9. Inquiry

   (a) A request for an item of equipment

   (b) A request concerning the status of a piece of equipment

   (c) A remote terminal

   (d) A communication with the computer asking for information of same type

10. TCTP

   (a) An entry on a computer card

   (b) Name for a computer file at Hq ATC

   (c) Orders from higher headquarters to perform a modification

   (d) Name for a computer file at Hq AFSC

11. Excess

   (a) Term applied to an item due for cannibalization

   (b) An item on hand, but which is not required

   (c) Code used to enter an item into a computer file

   (d) Term applied to delivered items that were not ordered

12. Demand level

   (a) Stock control level based upon past demands

   (b) Stock control level based upon current demands

   (c) Stock control level based upon future demands

   (d) All of the above

13. MILSTRIP

   (a) Military Storage Reports

   (b) A detailed listing of specifications used to describe items

   (c) Computer file

   (d) Procedures used to requisition items from source of supply

14. RDO

(a) Requisition directed order

(b) A group of regulations dealing with shipments

(c) Requested due-out

(d) A cause for a directed shipment

15. Quality Control

(a) Examining documents to insure they meet authorized procedures

(b) Examining equipment items to determine their physical condition

(c) Control of the stock level

(d) None of the above

16. ASC

(a) Code used to determine the location of an item

(b) Code used to indicate condition of an item

(c) Code used to authorize equipment to an activity

(d) Acquisition supply control

17. CA/CRL

(a) Authorized/in-use detail record

(b) Custodian Request/Receipt Listing

(c) A listing of all equipment items issued to a custodian.

(d) A listing of all equipment items authorized and in-use for each property custodian.

18. Status notices

(a) Notices used to relate action taken on a requisition

(b) Notices used to indicate the quality of an item

(c) Notices used to indicate the quantity of an item on hand

(d) A communication with higher headquarters

19. Special Levels

(a) Consist of minimum, maximum, and fixed levels

(b) Consist of predetermined and base initiated levels

(c) Consist of additive and adjusted levels

(d) All of the above

20. MEMI

(a) A single publication listing all equipment items

(b) Master equipment managed items

(c) Listing of parts that go together to make up a major assembly

(d) None of the above

21. Polling

(a) A partial inventory of equipment items

(b) A process for communicating with the computer

(c) When the remote is in-line with the computer

(d) A clicking sound

22. EMO

(a) A button on a terminal

(b) Code indicating the condition of an item

(c) An authorization code

(d) None of the above

23. Recurring Demand

(a) An item which has been obtained from supply and then returned

(b) A special computer request for information

(c) Under the stock fund concept, it refers to the sale of an item already sold once.

(d) A request for replacement of items used in normal day-to-day operations.

24. Cycle Inventories

(a) Conducted upon request

(b) Conducted every 3 months

(c) Refers to an item that was once new, but which has worn out.

(d) Conducted when account is transferred.

25. LCL

(a) Logistics control

(b) A button on a remote keyboard printer

(c) A listing of equipment items

(d) None of the above.

26. Due-out

(a) An item ready for cannibalization

(b) An item ready for shipment

(c) What occurs after due-in

(d) None of the above.

27. Allowance

(a) The maximum quantity of an item required to do a job.

(b) Number of equipment items authorized a given organization

(c) Refers to the equipment budget of an organization

(d) A special authorization

28. Nonlisted items

(a) Items which have not yet been inventoried

(b) A new shipment of equipment items

(c) Items which do not have a federal stock number

(d) Items which have not been accounted for

29. IPB

(a) Bench stock

(b) A list of items that go together to make a major assembly

(c) A management office

(d) Orders authorizing an equipment shipment

30. Transfer

(a) A transaction in which items are sent from one base supply warehouse to another.

(b) Refers to interchangeable items

(c) A special authorization from higher headquarters

(d) None of the above

31. FSN

(a) A special grouping of parts

(b) A transaction identification code

(c) Federal supply class

(d) An authorization code

32. Inventory Count Card

(a) An inventory overage document

(b) Lists the status of all equipment items on inventory

(c) Used to tag all equipment items during inventory

(d) Used to record physical quantity found in the custodian's account.

33. Suffix (of an AFSC)

(a) An entire AF career field

(b) A computer file

(c) Same as shredout

(d) None of the above

34. Interchangeable item

(a) An item which can be used for several purposes

(b) An item that can be used in place of another item with little or no modification.

(c) An item under the stock fund concept

(d) Same as substitution item

35. Nomenclature

(a) A full description of an item

(b) Specification of an item IAW ATCR 70-23

(c) Refers to a procedure for obtaining authorized items

(d) Refers to a special type of inventory.

Directions Mark only one answer to each question. Indicate how "sure" you are of your answer by checking one of the blocks.

	Very Sure	Sure	Fairly Sure	Not Very Sure	Not Sure
1. ___	[ ]	[ ]	[ ]	[ ]	[ ]
2. ___	[ ]	[ ]	[ ]	[ ]	[ ]
3. ___	[ ]	[ ]	[ ]	[ ]	[ ]
4. ___	[ ]	[ ]	[ ]	[ ]	[ ]
5. ___	[ ]	[ ]	[ ]	[ ]	[ ]
6. ___	[ ]	[ ]	[ ]	[ ]	[ ]
7. ___	[ ]	[ ]	[ ]	[ ]	[ ]
8. ___	[ ]	[ ]	[ ]	[ ]	[ ]
9. ___	[ ]	[ ]	[ ]	[ ]	[ ]
10. ___	[ ]	[ ]	[ ]	[ ]	[ ]
11. ___	[ ]	[ ]	[ ]	[ ]	[ ]
12. ___	[ ]	[ ]	[ ]	[ ]	[ ]
13. ___	[ ]	[ ]	[ ]	[ ]	[ ]
14. ___	[ ]	[ ]	[ ]	[ ]	[ ]
15. ___	[ ]	[ ]	[ ]	[ ]	[ ]
16. ___	[ ]	[ ]	[ ]	[ ]	[ ]
17. ___	[ ]	[ ]	[ ]	[ ]	[ ]
18. ___	[ ]	[ ]	[ ]	[ ]	[ ]
19. ___	[ ]	[ ]	[ ]	[ ]	[ ]
20. ___	[ ]	[ ]	[ ]	[ ]	[ ]
21. ___	[ ]	[ ]	[ ]	[ ]	[ ]
22. ___	[ ]	[ ]	[ ]	[ ]	[ ]
23. ___	[ ]	[ ]	[ ]	[ ]	[ ]

	Very Sure	Sure	Fairly Sure	Not Very Sure	Not Sure
24. ___	[ ]	[ ]	[ ]	[ ]	[ ]
25. ___	[ ]	[ ]	[ ]	[ ]	[ ]
26. ___	[ ]	[ ]	[ ]	[ ]	[ ]
27. ___	[ ]	[ ]	[ ]	[ ]	[ ]
28. ___	[ ]	[ ]	[ ]	[ ]	[ ]
29. ___	[ ]	[ ]	[ ]	[ ]	[ ]
30. ___	[ ]	[ ]	[ ]	[ ]	[ ]
31. ___	[ ]	[ ]	[ ]	[ ]	[ ]
32. ___	[ ]	[ ]	[ ]	[ ]	[ ]
33. ___	[ ]	[ ]	[ ]	[ ]	[ ]
34. ___	[ ]	[ ]	[ ]	[ ]	[ ]
35. ___	[ ]	[ ]	[ ]	[ ]	[ ]

Uniformed Physicians without Patients: A Squandering of Scarce Resources? [1,2]

Gary B. Brumback

American Institutes for Research

Washington, D.C. Office

In a country in which there is a doctor shortage, is it not a squandering of scarce resources when nearly 30,000 active physicians treat few patients or no patients at all? [3] Perhaps not. Some non-clinical functions may actually require physicians, and some physicians may actually seek out those functions. Optimal management of medical manpower would include keeping to a minimum the number of such functions and the number of physicians inappropriately assigned to them.

In this paper I shall focus on uniformed physicians in the administrative function. I have two reasons for so limiting my topic. I wish to share with you unpublished findings from an analysis I undertook especially for this paper of some data collected several years ago on professionals who were administrators in the Commissioned Corps of the United States Public Health Service (PHS). Secondly, the place of physicians in health care administration is becoming an increasingly prominent and controversial issue, and I believe it deserves more of our attention as researchers.

To give you some idea of how much opinions differ on the issue, I should like to quote five statements selected from the literature.

> "It scares me when I think of doctors getting into management roles. They do not have time, and I think they are not very good at it anyway. Further, I do not believe you can even train them to be very good at it (Guest, 1972, p. 311)."

You would be mistaken if you thought the above statement was made by a hospital administrator who had had one too many run-ins with the medical staff.

The person who made it was a professor of organizational behavior at the time, and he was participating in a discussion of his book chapter on "the role of the doctor in institutional management."

Consider a second quote. The president of a private philanthrophy recently said:

> "The kind of physician envisioned in this [new] program [of simultaneously training medical residents in management, economics, and other non-clinical fields] is seen by the Foundation as crucial to national progress in health (Rogers, 1973, p. 4)."

In its report to a former Secretary of the U. S. Department of Health, Education and Welfare, the Advisory Committee on Hospital Effectiveness made the following statement:

> "---appropriate numbers of the medical staff shall be directly involved with the administration in developing the budget and operating plan, and in the achievement of financial and service objectives as budgeted and planned--(HEW, 1968, p.)."

The fourth and fifth views are quoted from a Federal agency recommendation and a reply by the Federal agency under investigation:

> "We recommend [consideration be given to] assigning nonmedical personnel to administrative and management positions at the headquarters levels [and to beginning] a program to place MSC officers with master's degrees in hospital administration into positions as hospital commanders (The Comptroller General of the United States, 1971, p. 21)."

> "In the military,---the commanding officer of the hospital---has command, military, judicial, and professional responsibility as well as administrative jurisdiction over his staff---[and] must be a physician (The Comptroller General of the United States, 1971, p. 41)."

You may recognize the fourth view as a recommendation from the General Accounting Office following its 1971 study of medical personnel in the military, and you may recall that the reply came from one of its three services.

I might add that of the hortatory literature on the issue which I have reviewed so far, opinion seems to be running well in favor of giving more

physicians more administrative or management responsibilities. There is also the view by some that the resolution of the Nation's health care crisis requires more, not less medical leadership.

Having already introduced to you the specific subject of my paper, its course can be outlined thusly. First, I shall review the military utilization of physicians in administrative roles. Second, I shall summarize my review of the literature. Third, I shall summarize my own research findings on the subject. Finally, I shall suggest some research needs.

Physician Administrators in the Military

My review of the military situation will be brief and qualified because I have had only an outside perspective. I cannot even say that I am an outsider looking in for I have never observed first-hand the work of physicians in the military. Moreover, my look has been too quick for me to gain anything more than a general impression.

I have studied the staffing pattern for military physicians and for physicians in other Federal departments over a five-year period, 1968 to 1972, by examining end-of-the-year statistical data provided to me by the Department of Survey Research of the American Medical Association. These data are compiled by the AMA from their Physicians' Professional Activities questionnaire which is mailed out to all physicians. These data can presumably be accepted as reliable for I noted that the number of physicians in command and administrative positions that was quoted in a June 1973, Department of Defense news release was identical to the number tabulated by the AMA.

While there are more administrative physicians in the DoD than in any other single Federal agency, the percentage of the total medical manpower assigned to the administrative function the DoD is the second lowest among all of the agencies. The military staffing pattern for administrative and command positions has been relatively level over the five-year period studied, although there is a slight downward trend which began in 1970 (see Figure 1).

The mean five-year percentage of the medical manpower in the administrative function is slightly over seven and one-half (7.67%). In 1968, eight percent of the military physicians was in administrative and command positions, and by the end of 1972 the percentage had dropped to about seven. There has also been a similar downward trend among the other large Federal agencies.

A number of blue ribbon reviews, investigations, and analyses have been made over the years of military management of medical manpower. In 1969, Lt. Col. Gilbert Jacox of the Medical Service Corps reviewed 13 intra and extramural studies of factors affecting the retention of medical officers (Jacox, 1969). He identified and ranked 14 factors according to his assessment of their negative effect on retention. Three of these factors are more relevant here and illustrate very well how complicated the personnel management of professionals can be.

Consider the instability due to frequent rotations of duty assignments. Rotation is an equitable means of staffing hardship posts, it is necessary when medical officers are selected for training programs, and it may be an effective way of preparing young officers for command positions. On the other hand, rotation is the greatest deterrent to a medical career according to Jacox. Further, as Lando (1970) has pointed out in his study for the 1970 President's Commission on an All-Volunteer Armed Force, rotation disrupts the relationship between physician and patient and the teamwork needed in surgery.

Another factor is the assignment of medical officers to administrative and command positions. Three of the 13 studies reviewed by Jacox concluded that this factor was a major one in the turnover of medical officers, and Jacox himself ranked this factor eighth among the 14 in contributing to turnover. Jacox noted the apparent paradox of the medical leadership arguing for the necessity of physician administrators and commanders while at the same time separating medical officers were citing their dread of such assignments as a motivation for leaving the military.

Although not regarded by Jacox as having too significant an impact upon retention, a third factor which is relevant here is that of the professional and Departmental leadership of military physicians. Two of the 13 studies Jacox reviewed concluded that lack of professional leadership was a major obstacle to retention of medical officers. It was further concluded that reliance upon the draft as a recruitment source of physicians permitted a more casual management of physicians and medical programs. If this conclusion is valid, we should expect management style to change with the recent abolition of the draft. In any event, we may continue to have what I believe is still another paradox. There is the general thinking perhaps that one of the reasons it is necessary for a health administrator or manager to be a physician is the professional respect he automatically enjoys from the physicians under him. Unless the physician is a top notch administrator, or has a strong administrative deputy or aid, he is under the strain of attempting to achieve respect as a leader as well. Yet if he needs a strong deputy, then he is vulnerable to criticism for the very problem we began with, the staffing of administrative and command positions with physicians.

This problem was investigated in 1971 by the General Accounting Office in the study (The Comptroller General of the United States, 1971) mentioned above. In this study a number of billet descriptions of administrative and staff medical officers' positions were examined, and the officers were asked if their work required a medical background. All but a few officers believed their medical background was necessary. Nevertheless the investigators found what they believed were a number of questionable staffing practices, such as administrator-deputy pairs in which both incumbents were physicians. The investigators concluded that officers other than physicians could be placed in charge of medical facilities and in administrative and management positions at the Departmental level.

One year ago, the Secretary of Defense and the Secretary of Health, Education and Welfare issued a joint report (DoD-HEW, 1972) to the President and Congress on ways of reducing the reliance on medical officers while also meeting the medical needs of the armed forces. The report concluded that a reduction DoD manpower requirements for medical officers was possible (the projected reduction amounted to about a 20% shaving from the authorized strength of June 30, 1972). The report summarized five ways in which the reduction could be effected. While substitution of non physicians in the administrative function was not one of the five mechanisms mentioned, the report did note that an experimental test of the concept of small treatment facilities under the command of non physicians was then being made.

Apparently this test was at least partially effective. As you may know, the Deputy Secretary of Defense announced on June 1 of this year that there will be a gradual implementation of a policy which permits the placing of qualified officers without consideration of their basic professional status in positions of command or administrative direction of military health facilities. Thus such positions will no longer be filled only by physicians. Further, physicians performing administrative duties in staff level positions will be replaced by other health professionals as much as possible.

Given the reputed scarcity of physicians, their utilization within the DoD in work not directly related to patient care should continually be open to objective review. As you may know, the General Accounting Office is preparing to launch another and broader investigation into the management of health professionals in the military. The Office of Management and Budget, I am also told, will soon initiate a large-scale study of this general subject.

What questions could we ask ourselves today about how to manage a scarce resource which has growing and competing demands placed upon it? It seems a few very basic questions would be these: Does the work absolutely require a physician? If yes, who are the most qualified physicians for the work?

Should there be a definite administrative career path from entrance level to top command? If yes, what effect does this have on recruitment and subsequent retention of physicians?

I do not pretend to have the answers to these and similar questions. But before proceeding to the next section of this paper, I should like to offer the following general impression of mine. It appears to me the DoD may be caught up in a dilemma in a period of uncertainty. On the one hand, we have witnessed a slow decline in the percentage of medical officers in administrative roles over the last few years. This decline is perhaps in response to the very investigations and in-house analyses which I have reviewed here today. The decline has just become stated Departmental policy, and it could very well be an absolute necessity for an all-volunteer armed force. On the other hand, however, we are also witnessing growing opinion in the health field in favor of physicians becoming more active in the management of health care facilities. What is the solution? There is scarcely any evidence, as we shall see in a moment, to support or refute either argument. It is difficult to be very certain about management decisions in the face of such uncertainty.

Review of the Literature

As far as I can tell from my review of the literature on the issue of physicians in administrative roles, there has been quite a bit of opinion and no research. Almost all of the opinion, incidentally, appears in the literature of the health field, such as the journal of the American Hospital Association, Hospital. A MEDLARS search which I requested from the National Library of Medicine of articles indexed by MEDICUS netted 68 citations, not one of which was a report of any research study. I reviewed five and one-half years of Psychological Abstracts covering the period of January, 1968, to July, 1973. Not one of the approximately 1300 citations indexed under either the subject heading of "medicine" or "physician" represented a research study of the issue.[4] Nor did I find any research study out of more than 700 citations

during that same period under the heading of "personnel in the military." I also requested a Defense Documentation Center search for any published or unpublished documents on the issue. This search yielded nearly 190 citations, and not one was a report of any research on or an analysis of the issue.

Other than the liklihood that uncertainty rules over the issue, what else can be concluded from my literature review? First, since a considerable body of knowledge has been accumulating on leadership and managerial effectiveness (there are, for example, nearly 800 references listed in the highly regarded book by Campbell, Dunnette, Lawler & Weick, 1970, on managerial behavior), it is the medical profession and health care establishment, and not the administrative function itself, obviously, where researchers have failed to make an entry. Why is this so?

There is the belief that physicians are unwilling subjects because they have too little time to spare and/or they see research as a threat to the status quo. I question how close to the truth this belief is. Several thousand physicians in a large bureaucracy have willingly participated in my own research. I suspect the medical profession and health care establishment may be somewhat sensitive about the issue, but they are also perceptive enough to recognize that resistance to growing demands for public inquiry would be self defeating. Unfortunately, industrial and organizational psychology, one of the disciplines capable of assisting the medical profession and health care establishment to examine the issue objectively, apparently has either ignored or been unaware of the issue (I should add that psychologists in general have had a myopic view of health and health care; one of my goals on other occasions has been to promote a broader view of the problems which pervade the health care industry-e.g., Brumback, 1972).

My second conclusion is that until we have the research findings we need on specific questions about physicians as managers, we should begin

trying to draw parallels among administrative positions in medicine and health care and those positions in the larger segment of the managerial world where so much knowledge already exists.

Unpublished Research on PHS Physician Administrators

I wish now to turn briefly to my own unpublished research on physician administrators in the Commissioned Corps of the PHS. This research should be regarded as strictly exploratory because the bulk of data I have analyzed was a by-product of a multi-phased research program which I designed more than five years ago for a purpose other than testing specific hypotheses about physician administrators. Therefore, the findings I am presenting here are a result of post hoc analyses of a potpourri of data.

The research program was an effort to develop and validate a job-content based performance appraisal system for commissioned officers in the PHS. The program and its major results have been reported elsewhere (Brumback & Vincent, 1970a, 1970b; Brumback & Howell, 1971a, 1971b, 1972).[5] To give you some background information, the first phase of the program involved a factor analysis and then a hierarchical cluster analysis (a la Christal, 1962) of a standardized position inventory which had been constructed and administered to about 5000 PHS officers. The second phase involved collecting over 2,500 critical incident reports of job performance from about 500 officers. The third and final phase was a validation of an experimental efficiency report against a criterion of averaged work associate ratings on a sample of about 1200 officers.

I have explored five questions using some of the by-product data from the above program. For each question, the independent variable is the type of position held by the physician or his comparison group. The two types of positions of concern here are administrative positions and patient care positions. Position type was determined by the hierarchical clustering technique mentioned

above.[6] This technique had produced quite distinguishable position types or job families. For example, there were no patient care duties among the top 50% of the duties which accounted for the group job description of physicians in the administrative job family. I have limited my analyses to Director grade officers within each of the job families because officers at this next to the highest level are careerists. Further, there were too few Surgeon General grades for statistical comparisons.

My first question was this: were physicians who became career administrators less clinically proficient at the time of entry into the PHS than were physicians who remained as clinicians throughout their PHS career?

I was able to trace back to the entrance professional examination records for a physician sample of career administrators (N = 18) and a matched physician sample (N = 18) of career clinicians.[7] At that time, physicians were selected into the Corps on the basis of a final weighted average of their written and oral examination scores and file or professional background review scores (see Newman, 1951, for a description of the officer selection program of the PHS). I computed the mean of the final standard selection scores for each group and found that they were identical (a mean score of 84 for each group). Perhaps contrary to one's expectation, career administrators were not judged any less qualified as clinicians than were the career clinicians at the time of entry into the PHS.

My next question was: at the time of entry into the PHS, did career administrators have initially different career intentions and work preferences than did the career clinicians?

I was able to retrieve and examine for two matched samples of physicians (N = 56) their responses to a Service Evaluation Questionnaire (SEQ) which had been administered as an attitude survey to officers several years prior to my own research program (see Howell, Brumback, Newman & Rizzo, 1968, for a description of the SEQ research).[8] I found no difference between the two groups

in their initial career intentions (Chi Square = 1.1, $p > .05$). For both career administrators and career clinicians, only slightly over 10% of them stated they initially had planned to leave the PHS after fulfilling their military obligations. The majority of the physicians were either open-minded, favorably disposed, or committed to a career at the time of entry. I did find a significance difference (Chi Square = 36.7, $p < .001$) between the two groups in their initial work preferences. Whereas most (84%) of the career clinicians preferred patient care work at the outset, career administrators were about evenly split between preferring patient care work and various other kinds of initial assignments.

Thirdly, I was curious as to whether career administrators are less satisfied with their careers, which some might regard as marginal role playing, than are career clinicians. I found no mean difference ($t = 0.7$, $p > .05$) between overall SEQ scores for the two groups. Both groups were equally satisfied with the careers they had made.

Fourthly, I wanted to know whether high-level physician administrators in the PHS become administrators over night or have had previous administrative assignments. My results indicate the latter; physicians who became high-level administrators had worked their way up through a progression of increasingly responsible administrative positions. Physicians in the administrative job family had served an average of about nine and one-half years in administrative assignments before reaching their Director grade posts. Moreover, some of these physicians had acquired administrator training through graduate degree programs in public health.[9] In addition to progressing through a career path of administrative positions, the high-level administrators typically had rotated through a variety of non-administrative positions earlier in their career. Prior to their succession of administrative assignments, these physicians had served an average of about 10 years in non-administrative positions.[10]

My last question was this: how effectively do physicians perform as administrators compared to other professionals in administrative positions? This could be called the $64 question.

If physicians do not make very good administrators, as the professor of organizational behavior mentioned earlier fears, my findings indicate at least that physicians do not appear to be any worse administrators than other professionals in administrative positions. Although more critical incidents were reported on physician administrators than would be expected according to their representation in the PHS (Chi Square = 21.40, $p < .001$), there was no significant difference between the number of reported negative incidents of administrative performance of physicians and the number of negative incidents reported on other non-physician administrators (Chi Square = 3.77, $p > .05$). Further, analysis of variance showed no significant difference in the mean work associate ratings of effectiveness of administrative performance among physicians and 10 other professional groups ($F = 1.56$, $p > .05$). Considering the fact that the physicians tended to occupy the more responsible administrative positions, even with grade level constant, than did the other professional groups, physicians would appear to have more than held their own in the discharge of their responsibilities.[11]

I should like to draw some tentative conclusions from my research. First, physician administrators do not at the beginning of their career avoid or leave clinical medicine because they lack clinical proficiency. Indeed they were judged no less proficient as clinicians than were the career clinicians. Second, some physician administrators prefer non-clinical work at the beginning of their career even though they probably would have been proficient clinicians in practice. I suppose the reasons for this are to be found in the interests, personality and biographical characteristics of such physicians. Third, physician administrators are

not frustrated actors playing in marginal roles. The physician administrators I studied appeared to enjoy their work and career. Fourth, physicians do not make bad administrators. Fifth, physicians can be prepared and trained to be good administrators.

Further Research Needs

The utilization of physicians in administrative roles is an important and controversial issue which begs for more research attention. I should think that the $128 question that needs to be answered is whether or which health administration and management positions absolutely require a medical education. Do some of these positions really require physicians, or is it more a case of some physicians really requiring the positions? What is needed is a systematic classification or taxonomy of all higher level positions in the health and health care field across organizations and functions. This classification should be determined by the critical knowledge requirements of policy and decision making and the critical interpersonal skills needed. If policies, decisions and leadership do not demand a medical education, then a high level administrative or management position should not be filled by a physician. The classification should also distinguish those positions in which medical leadership is not critical, but in which close medical consultation to the leader is.

I wonder if the state-of-the-art of our job analysis methodology is up to the challenge. Obviously a physician incumbent's rationale for his assignment is not going to be a sufficiently valid justification for requiring a medical education. Comparisons of staffing patterns among organizational units may reveal inconsistencies, but these may be due just as much to differences in actual requirements as to arbitrary practices. One needs more data than the mere fact in two or more similar units one is headed by a physician with a certain position title and the other is headed by a non physician with the same title. One would also need more than the standard

position inventory analysis to determine critical knowledge requirements. For the moment I frankly am not sure how I would proceed methodologically. I suppose a possible approach would be to have both a job analyst and a subject matter expert thoroughly analyze the content of each critical decision and policy and each leadership act for a sample of administrative positions held by physicians. The information could then be used to build an in-basket test. If non physicians could reproduce the policies, decisions and leadership acts of the physicians, there would be more reason to believe that a medical education is a critical requirement.

Another important area for research, assuming there are indeed a number of administrative positions which require a medical education, is the early identification and development of managerial talent among physicians or even among medical students. What are the individual differences that can be measured and used to predict subsequent specialization in health administration? Biographical data as predictors should be explored as well as interest and personality measures. Howell and Newman (1967) reported on a concurrent validity study of a combination biographical and personality inventory. These researchers were able to differentiate the most clearly between physicians in medical care and public health-administration on the basis of answers to the inventory.

Another crucial question is when and how during the process of medical education or professional practice should physicians learn administrative and managerial skills? A new program for giving management training to medical residents was cited above. What should be the content of such a program? As Campbell et al. (1970) have observed, there has been little research on training content as opposed to techniques in management development and training programs. Longitudinal research is also needed on the kinds of career experiences which lead to successful administrative assignments.

References

Brumback, G. B. (Chm) Toward a psychology more responsive to the public health. Symposium presented at the annual meeting of the American Psychological Association, Honolulu, September, 1972.

Brumback, G. B. & Vincent, J. W. Factor analysis of work-performed data for a sample of administrative, professional and scientific positions. Personnel Psychology, 1970, 23, 101-107. a

Brumback, G. B. & Vincent, J. W. Jobs and appraisal of performance. Personnel Administration, 1970, 33, 26-30. b

Brumback, G. B. & Howell, M. A. Rating clinical performance. Journal of Dental Education, 1971, 35, 43-46. a

Brumback, G. B. & Howell, M. A. Evaluating the professional performance of staff pharmacists: A methodological study. American Journal Hospital Pharmacy, 1971, 28, 336-342. b

Brumback, G. B. & Howell, M. A. Rating the clinical effectiveness of of employed physicians. Journal of Applied Psychology, 1972, 56, 241-244.

Campbell, J. P., Dunnette, M. D., Lawler, E. L. III., & Weick, K. E. Jr. Managerial behavior, performance, and effectiveness. New York: McGraw-Hill, 1970.

Christal, R. E. USAF occupational research projects of tri-service interest. In Proceedings: Tri-service conference on new approaches to personnel systems research. Washington: Office of Naval Research, ONR Symposium Report ACR-76, 1962, 62-75.

Guest, R. H. The role of the doctor in institutional management. In B. S. Georgopoulos (Ed.), Organization research on health institutions. The University of Michigan: Institute for Social Research, 1972.

Jacox, G. L. A compendium of studies on career retention: Factors relating to early resignation or retirement of physicians in the military medical service. Unpublished report. Department of Defense, July, 1969.

Lando, M. Health services in the all-volunteer armed force. In Studies prepared for the President's Commission on an All-Volunteer Armed Force. Washington, D. C. November, 1970.

Newman, S. H. The officer selection and evaluation program of the U. S. Public Health Service. American Journal of Public Health, 1951, 41, 1395-1402.

Newman, S. H. & Howell, M. A. Prediction of work area specialization among professional health personnel. Personnel Psychology, 1967, 20, 89-110.

Howell, M. A., Brumback, G. B., Newman, S. H. & Rizzo, J. R. Work satisfaction and the retention of organizationally employed physicians. Personnel Psychology, 1968, 21, 63-77.

Rogers, D. E. quoted in the American Institutes for Research Behavioral Science Newsletter, 1973.

The Comptroller General of the United States. Improvements needed in establishing requirements for, and uses of, medical professional personnel in the military services. Washington, D. C. December 16, 1971.

The Secretary of Defense and the Secretary of Health, Education and Welfare. Reducing the needs for military medical personnel in the armed forces. Washington, D. C. October, 1972.

The United States Department of Health, Education and Welfare. Report of the Secretary's Advisory Committee on Hospital Effectiveness. Washington, D. C. 1968.

Footnotes

[1]Paper presented at the 1973 Military Testing Association Conference, San Antonio, Texas, 28 October -- 2 November 1973.

[2]The writer is indebted to a number of persons for their contributions at various stages of the research and preparation of this paper. I wish to thank in particular two staff members of American Institutes for Research, Mrs. Roberta Marcum for assistance in the data analyses and Mrs. Lily Griner for library support in the literature review. I also acknowledge with thanks Dr. Jimmy L. Hatfield of AIR for his consultation on information sources. I am also deeply grateful to Miss Gene Roback and her staff of the Department of Survey Research of the American Medical Association for providing me with statistical data on the distribution of physicians in the United States over a five-year period. Thanks are due several Department of Defense individuals, Mr. Mark Biegle, Dr. Ralph Canter, Col. Wm. S. Augerson and Col. Billy F. Simmons for their helpful information obtained via telecons. In addition, Col. Simmons was especially helpful in providing me with documented materials. Thanks are also due several General Accounting Office individuals, Mr. Charles Collins, Mr. Gene Hammel and Mr. Max Stettner for their helpful information. Impossible though it would have been to complete the research and paper with out the contributions of the above named persons, they are in no way responsible for whatever imperfections may appear in this presentation.

[3]This rounded figure indicates the approximate number of active and classifiable physicians in the United States whose primary professional activity is not directly involved in patient care. The figure is based on data compiled from a computer listing furnished to me by the Department of Survey Research of the American Medical Association. The AMA criterion of primary activity is the largest percentage of time spent by the physician in a single activity.

The AMA defines direct involvment in patient care as the seeing of patients, but also includes the patient services of such physicians as pathologists and radiologists.

[4]The total count of 1300 citations is somewhat inflated since there are some overlapping citations under the two headings of "physician" and "medicine."

[5]Reprints of published work from this research program or from related efforts are available from the writer upon request.

[6]Position type was identified from the hierarchical clustering technique for all but the last question. In that instance, administrative physicians were identified by their scores on work associates' ratings of functional performance and by type of critical incident report.

[7]The two samples were matched on entry-on-duty date.

[8]The two samples were matched on tenure.

[9]Fifty-five percent of the physician administrators had graduate degrees. It is not known what percentage of this group had master or doctoral degrees in public health administration.

[10]It is not known what the assignment patterns were for the physicians before they became administrators. Presumably some of them began their PHS employment as clinicians.

[11]Physicians supervised significantly more staff and controlled larger budgets than did the other professional groups.

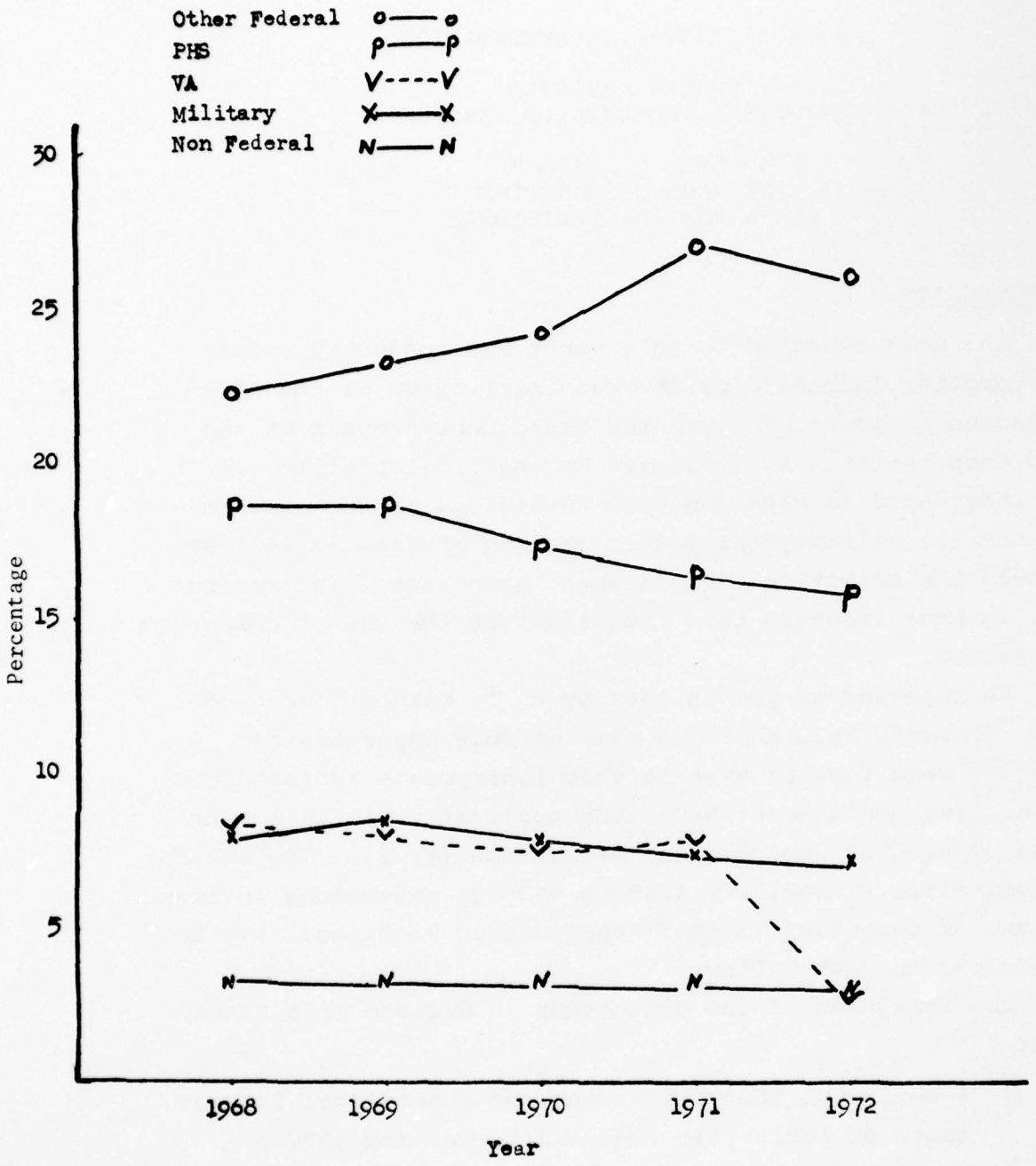
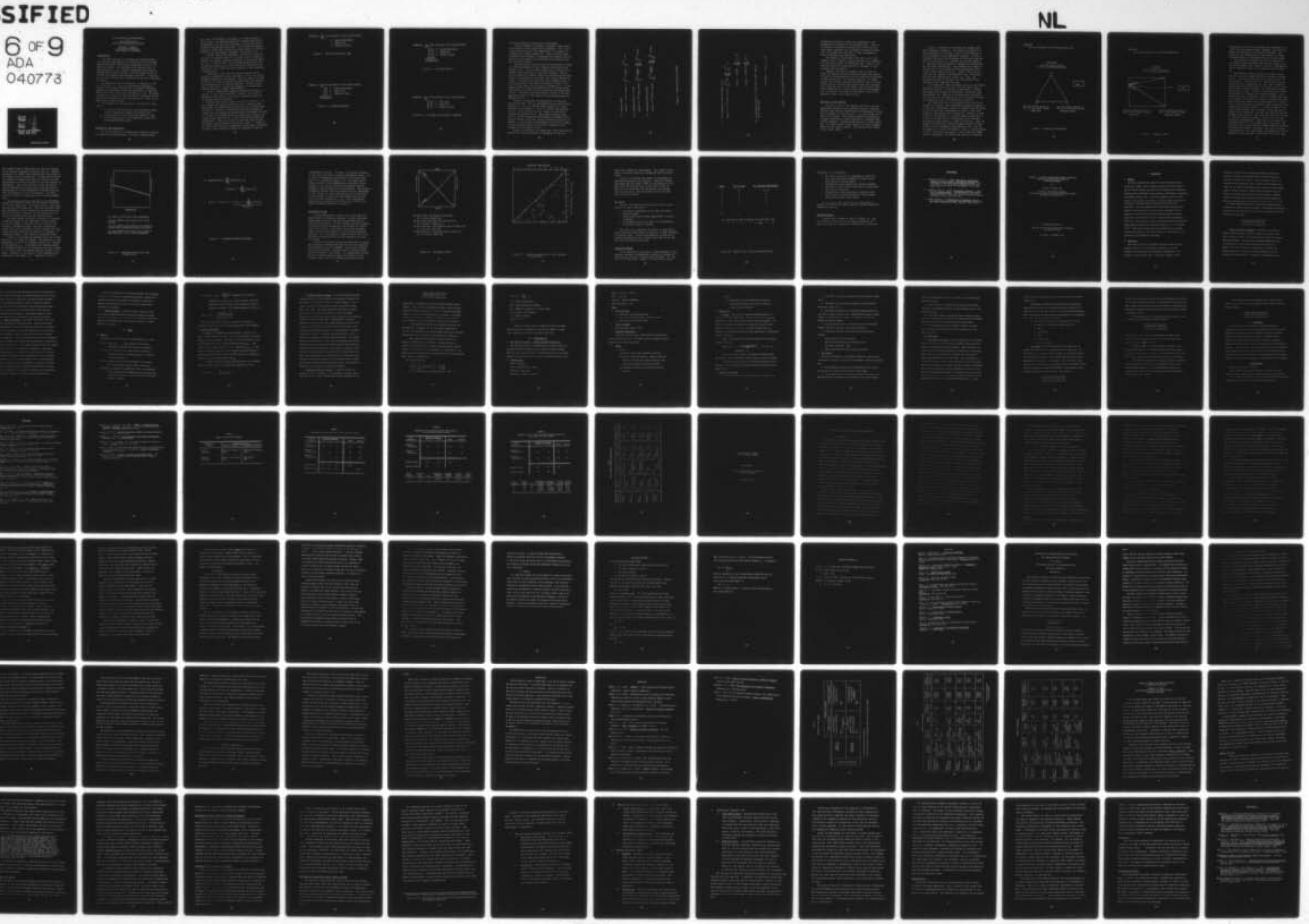


Fig. 1. Percentage of active and classifiable physicians in the USA whose primary professional activity is administrative work. (Based on data compiled by the Department of Survey Research of the American Medical Association).

AD-A040 773 AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TEX F/G 5/9
PROCEEDINGS, ANNUAL CONFERENCE OF THE MILITARY TESTING ASSOCIAT--ETC(U)
NOV 73

UNCLASSIFIED NL

6 OF 9
ADA
040773



AN EXPERIMENTAL IMPLEMENTATION OF COMPUTER ASSISTED ADMISSIBLE PROBABILITY TESTING

WILLIAM L. SIBLEY
THE RAND CORPORATION
SANTA MONICA, CALIFORNIA

INTRODUCTION

The work reported in this paper was conducted under the Computer Technologies in Training Project of the ARPA-sponsored Manpower and Training Management Program of the Rand Corporation. As indicated by the project title, we are interested in ways computer technology can be used to enhance the military services' training program as well as to help the selection and placement processes. In particular, we have recently been investigating the use of computers in testing.

As reported to you in 1969 by W. C. Gardner, Jr. and E. H. Shuford, Jr., there is considerable experimental evidence that testing systems that incorporate admissible probability measurement have wide application in important areas of military testing and evaluation [1, 2]. The nature of Admissible Probability Testing and its extensions invites the use of computers in their application as I shall try to make clear in what follows.

The remainder of the discussion is divided into three parts:

- o A review of the motivations and theoretical foundations of Admissible Probability Testing (APT).
- o A description of the role of the computer in APT.
- o A brief discussion of our experience to date with the computer-based system.

MOTIVATIONS AND FOUNDATIONS

The prime motivation for Admissible Probability Testing is simple and straightforward: we need to measure better

the "state of knowledge" of trainees in a large variety of subject areas. An improved measuring technique would be of considerable value not only in testing a trainee but also in evaluating and planning curricula. As indicated by Gardner, planning is especially critical for those courses of instruction that are individually tailored to the trainee [1].

A second, and no less important, motivation is that the estimation of probabilities is a critical skill in some endeavors; for example, weather prediction, intelligence analysis, etc. I believe that it will become apparent that APT is especially applicable in the training of students to use effectively all of their facts and reasons in the estimation of probabilities.

A corollary of the need for training specialists in probability estimation is the idea that, in fact, we are all engaged in probability estimation. That is, there is a need for us all to convey accurately and effectively our opinions of the possibilities of various outcomes. If that is the case, we must learn to speak the language of probabilities precisely and with a common understanding.

So what is Admissible Probability Testing and how is it applied? I shall talk about APT in terms of our specific application even though it has greater generality than may be apparent from my discussion.

Consider the multiple choice test item in Figure 1. We are generally required to respond to such a question with what amounts to a probability distribution as is illustrated in Figure 2. However, it would be my personal tendency to respond with the probability distribution illustrated in Figure 3. In fact, if confronted with the question in Figure 4, my response might very well be, "I can't distinguish among the alternatives." That is, as far as I can tell from my facts and reasons, each alternative is equally likely. I think it is evident that requiring a response of zeroes and a single one can mask a large variety of states of knowledge as well as

QUESTION: The 11th president of the United States was:

1. George Washington
2. James Polk
3. Franklin Pierce

Figure 1. Multiple Choice Test Item

QUESTION: The 11th president of the United States was:

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

1. George Washington
2. James Polk
3. Franklin Pierce

Probability Distribution

Figure 2. A Standard Response

QUESTION: The 11th president of the United States was:

0	1.	George Washington
.5	2.	James Polk
.5	3.	Franklin Pierce

Revised Probability Distribution

Figure 3. A Revised Response

QUESTION: The 11th president of the United States was:

.3	1.	John Tyler
.3	2.	James Polk
.3	3.	Franklin Pierce

Figure 4. An "unable to distinguish" Response

introduce guessing on the part of the student.

If such freedom of response is allowed, there remains the problem of how to score such a response. That is precisely the function of the "admissible," or "proper," or "reproducing" scoring system [3]. Such a scoring system encourages the strategy of responding with exactly what you perceive the probability distribution to be based on your own facts and reasons.

The admissible scoring system can be motivated in terms of a gambling example (a more complete discussion can be found in [3]). For the sake of simplicity, suppose that a student were faced with a true-false question. Then he may be viewed as a gambler faced with determining the probability of occurrence of the event "the correct response is 'true'." Suppose also that there were a gambling house in which there were $\phi(u)du$ wagers available at odds $(1-u)/u$ (the "correct odds" for an event of probability u). If the student believes that the probability of the event taking place were p, he would take all wagers at odds better than $(1-p)/p$. Similarly, he would take all wagers on the event <u>not</u> taking place at odds better than those appropriate for probability 1-p.

Figure 5 illustrates the mathematics of admissible scoring systems. Line (1) corresponds to our true-or-false discussion. Line (2) extends those notions to a n-foil multiple choice question. Line (3) adjusts the equations to give a score of zero for the "unable to distinguish" response of $p_i=\frac{1}{n}$ for i=1,n. What these integrals represent is the net settlement made between the gambling house and the student, after the correct alternative has been revealed. It is the sum of settlements on an infinite number of infinitesimal wagers, each of which appeared to the student to be favorable to him.

Figure 6 illustrates the logarithmic admissible scoring system which has the property of depending on only the

(1) True or False

$$\text{payoff (if "true" occurs)} = \int_0^p \phi(u)\frac{u}{1-u}du - \int_0^{1-p} \phi(u)du$$

(2) Multiple choice (n foils)

$$\text{payoff (if } i^{th} \text{ event occurs)} = \int_0^{p_i} \frac{\phi(u)du}{u} - \sum_{j=1}^{n} \int_0^{p_j} \phi(u)du$$

(3) Adjusted Multiple Choice

$$\text{payoff (if } i^{th} \text{ event occurs)} = \int_{1/n}^{p_i} \frac{\phi(u)du}{u} - \sum_{j=1}^{n} \int_{1/n}^{p_j} \phi(u)du$$

Figure 5. Admissible Scoring Systems

(1) The Logarithmic Scoring System $\phi(u)=1$:

$$\text{payoff (if } i^{th} \text{ event occurs)} = \int_{1/n}^{p_i} \frac{du}{u} - \sum_{j=1}^{n} \int_{1/n}^{p_j} du$$

$$= \log(p_i) - \log(\tfrac{1}{n}) - \sum_{j=1}^{n} (p_j - \tfrac{1}{n})$$

$$= \log(np_i) \quad (\text{assuming } \sum_{j=1}^{n} p_j = 1)$$

(2) An adjusted system for 3 foils and a score range of about 100 points:

$$= 50 \log(3p_i) \qquad .01 \leq p_i \leq .98$$

Figure 6. The Logarithmic Scoring System

probability ascribed to each foil independent of the probabilities ascribed to the remaining foils. Line (2) of Figure 6 illustrates the particular version of the logarithmic system we use in Admissible Probability Testing. No probability less than .01 may be assigned to avoid infinitely large penalties.

The gambling illustration shows that it is in the student's best interest to gauge accurately his probabilities and thus to make no unnecessary bets. That is, the scoring system does encourage him to reproduce his own probabilities accurately.

However, how does the student learn to "play the game" and improve the way he brings his facts and reasons to bear on the determination of the probability he ascribes to each alternative? As Gardner points out, "It seems likely that the more exposure the students have to [probability] tests, the more realistically they will evaluate test items and their own knowledge, thus providing a more valid indication as to their actual level of achievement" [1]. We believe the answer to the question lies in the computer system I will describe next.

THE ROLE OF THE COMPUTER

As you will see in the figures that follow, the computer-based system we have developed requires a computer terminal with graphics capabilities and some means of allowing the student to interact with it in elapsed times of at most a few seconds. The original system was developed on a highly interactive console connected to a powerful IBM 370/155 computer. A newer system exists on a "smart terminal," the IMLAC PDS-ID, which supports the system independently of any larger computer. My discussion will relate to the latter system.

Figure 7 illustrates the display the student sees while answering a question. He assigns a probability to each alternative by choosing a point in the interior or on the boundary of the equilateral triangle. In such a triangle, the sum of the distances of the point from each side of the triangle is a constant and can be scaled to 1. Thus each point determines a probability distribution for the three alternatives. The probabilities appear at the vertices of the triangle and the appropriate logarithmic score appears close by. Figure 7 shows the results for the "unable to distinguish" case. The probabilities are constrained to sum to one. Also, the probabilities must be greater than .01 because of the behavior of the logarithm near zero.

The student may choose different points in the triangle until he is satisfied with the indicated probability or score distribution. The nearer he approaches a vertex, the nearer that probability approaches .98 and the other two approach .01. When he is satisfied with the distribution he selects the NEXT option and the terminal displays his cumulative score as illustrated by Figure 8.

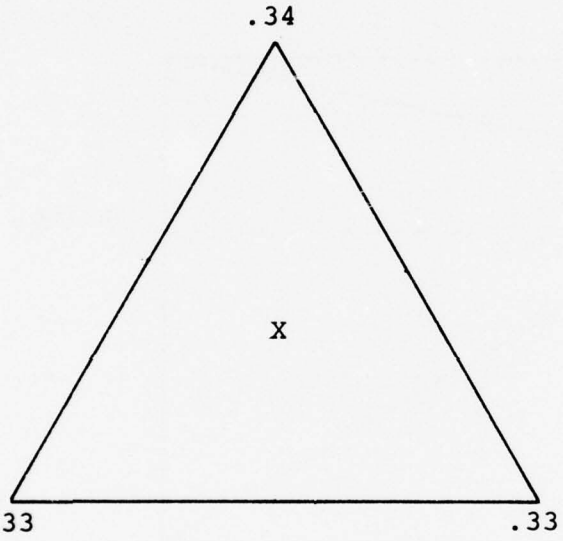
The cumulative score is presented in the form of the maximum (or minimum) score attainable question by question. In the case illustrated in Figure 8, nothing has been gained or lost by the score assignments in Figure 7 (the final segment in the broken line in Figure 8). However, a probability of .98 assigned to a correct answer would result in a gain of 23 points while .98 assigned to an incorrect alternative would result in a loss of 76 points. Thus, the student has question by question feedback about the relationship of his probability assignments to his score and may adjust his actions accordingly. That is, he may develop a "feel" for the scoring system. Under usual test administration procedures, the delay from testing to reporting

QUESTION:

The 11th president of the United States was:

JOHN TYLER

You will GAIN 0 points
if this is the correct answer.



NEXT

You will GAIN 0 points if
this is the correct answer.

JAMES POLK

You will GAIN 0 points if
this is the correct answer.

FRANKLIN PIERCE

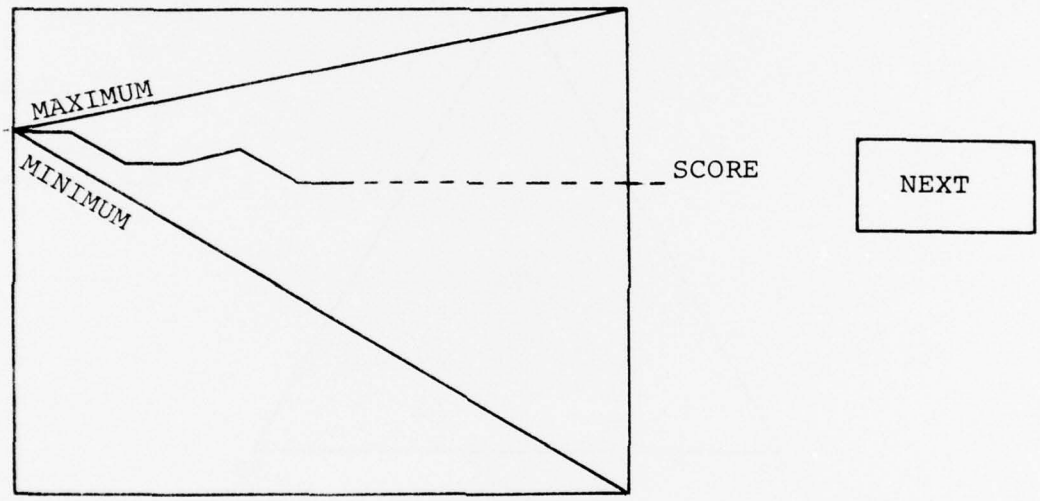
Figure 7. Probability Assignment

QUESTION:

The 11th president of the United States was:

JOHN TYLER

You will GAIN 0 points if
this is the correct answer.



You will GAIN 0 points if
this is the correct answer.

JAMES POLK

You will GAIN 0 points if
this is the correct answer.

FRANKLIN PIERCE

Figure 8. Cumulative Score

the scores is so long as to prevent the development of a "feel" for the scoring system. Moreover, the scores are usually not reported on an item by item basis. This portion of the computer-based system gives the student a maximum amount of experience with the scoring system.

Even though the student may learn to express his own probabilities accurately the question remains as to how those probabilities relate to reality. That question can be answered in part by his External Validity Graph [3].

The External Validity Graph can be used to determine how well the student does as an estimator of probability. It can be thought of as his "track record" as an estimator. We may collect all those events which he claims to have a probability of, say, .80 of occurring, and determine what percentage of them did, in fact, occur. If he were a perfect estimator, we would expect roughly 80% of them to occur. Figure 9 illustrates such a graph. The diagonal AB represents a perfect estimator. If we collect, say for 20 questions, all the probabilities used by a student and compute the relative frequency with which each particular probability is associated with a correct foil, we may construct an external validity graph for that student and that test material. We then assume that the relationship between them is roughly linear and apply a simple linear regression. We call the resulting line the student's realism line. The lines A'B' and A"B" in Figure 9 illustrate possible results. The line A'B' has the following interpretation: about 60% of the events given a probability of 1.0 of occuring did in fact occur. We interpret this to mean that the student tends to over-value his facts and reasons. Similarly, the line A"B" can be interpreted as: the events given a probability of .60 (or greater) of occuring did in fact occur 100% of the time.

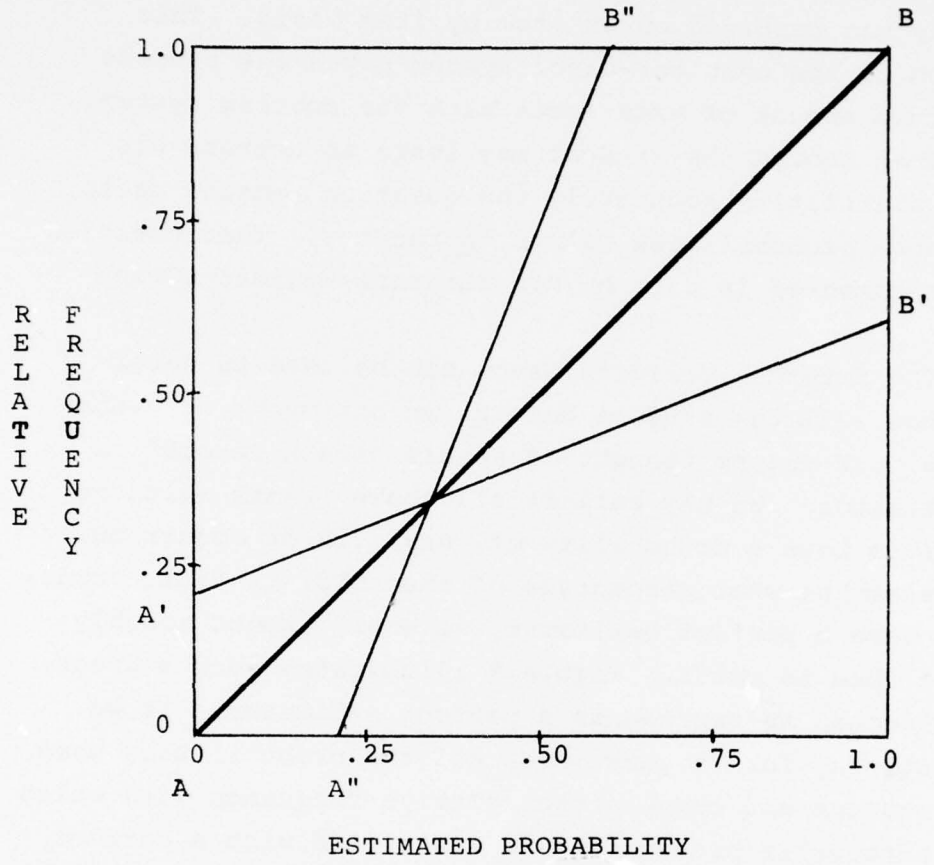


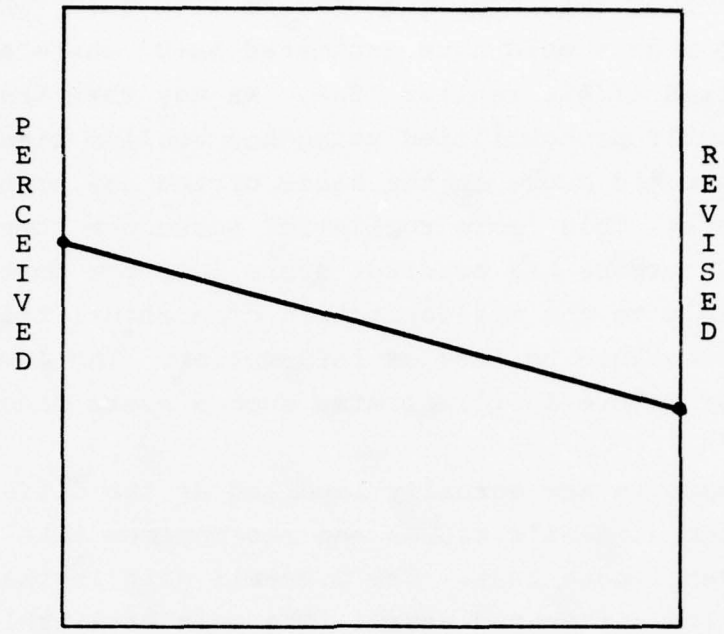
Figure 9. External Validity Graph and Realism Lines

The student tends to under-value his facts and reasons. We may further interpret the realism line as: "when he estimated x he should have estimated Ax+B" where Ax+B is the equation of his realism line. We may then transform the student's probabilities using his realism line and recalculate his score on the basis of the new probability assignments. This "more realistic" score can then be used to decompose his original score into the portion attributable to the mis-estimation of probabilities and that attributable to lack of information. The lower portion of Figure 10 illustrates such a score decomposition.

The scores are actually reported as the difference between the student's scores and the maximum attainable for a given length test. The over-all gain is that maximum less the un-revised score. The gain attainable from more information is that maximum less the revised score. The gain from more "realistic" use of probabilities is simply the revised score less the original score.

The upper portion of Figure 10 brings us to another interesting aspect of the logarithmic scoring system. We call the square in the upper part of Figure 10 the student's information square. The term "information" arises from a very interesting relation between the logarithmic scoring system and the concept of "entropy" in information theory. That relation is developed below.

Figure 11 illustrates the expected score for a given question in our three foil system. The expression in the parentheses is called in information theory the "entropy" of the partition p_i. It represents the expected amount of information which will be conveyed by the event itself; or, in other words, the expected surprise content of the event. Equation (2) of Figure 11 is the expected score normalized to the range 0 to 1. Equation (2) can be



YOU TEND TO OVER VALUE YOUR INFORMATION.

YOU CAN IMPROVE YOUR SCORE BY 261 POINTS OVERALL.

YOU CAN IMPROVE YOUR SCORE BY 236 POINTS BY GAINING MORE INFORMATION ABOUT THE SUBJECT.

YOU CAN IMPROVE YOUR SCORE BY 25 POINTS BY MORE REALISTIC USE OF YOUR INFORMATION.

Figure 10. Information Square and Score Decomposition

$$(1)\quad \text{Expected Score} = \sum_{i=1}^{3} p_i (50 \log 3\, p_i)$$

$$= 50 \left[\log 3 - \left(-\sum_{i=1}^{3} p_i \log p_i\right) \right]$$

$$(2)\quad \text{Relative information} = \frac{50 \left[\log 3 - \left(-\sum_{i=1}^{3} p_i \log p_i\right) \right]}{50 \log 3}$$

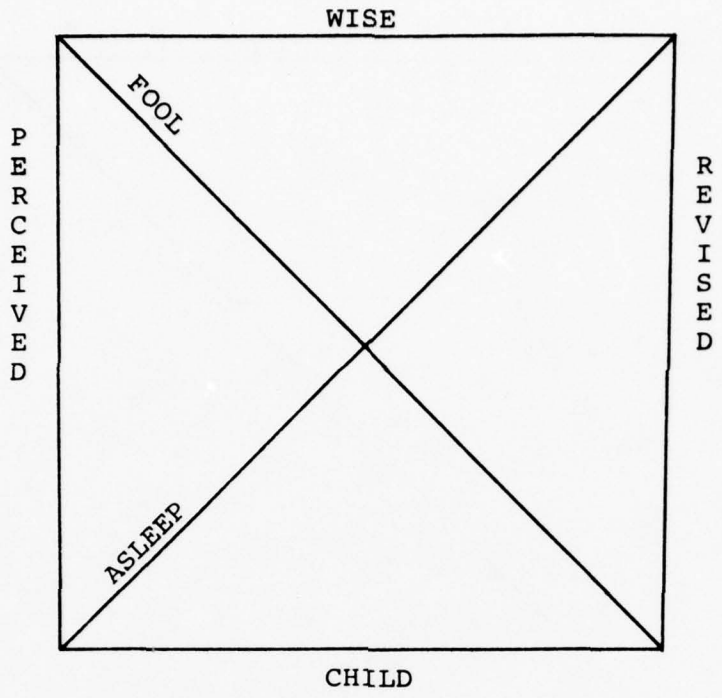
Figure 11. Information Theory Analogues

interpreted as follows: if there is no surprise content in the event then there is perfect information [eq(2) = 1]; if there is maximum surprise content in the event then there is no information [eq(2) = 0]. The upper part of Figure 10 plots on the PERCEIVED axis the point corresponding to the relative information computed from the student's original probability distributions. The relative information derived from the probabilities as transformed by the realism line (the regression line in the external validity graph) is plotted on the REVISED axis and the two points are connected by a line. This information square has an interesting interpretation in terms of an Arabian proverb as depicted in Figure 12.

EXPERIENCE TO DATE

The few observations I would like to make here are based largely on anecdotal evidence. We have used every opportunity to test our approach before committing to formal experimentation. We have gleaned our test students from among the Rand professional and secretarial staffs (and their children), high school and college students, casual visitors and attendees at various Rand-sponsored meetings. It is gratifying to report that not one of our subjects has had difficulty with the mechanics of using the system.

Figure 13 illustrates the relation between the actual and perceived information measures for the first exposure of 66 subjects to the test system. This figure supports what we have observed: subjects tend to overvalue their facts and reasons at the outset. Our observation has been that most subjects tend to choose the vertices of the triangle rather than try to differentiate more finely



He who knows, and knows that he knows,
He is wise, follow him.
He who knows and knows not that he knows,
He is asleep, awaken him.
He who knows not, and knows not that he knows not,
He is a fool, shun him.
He who knows not, and knows that he knows not,
He is a child, teach him.

Figure 12. An Arabian Proverb

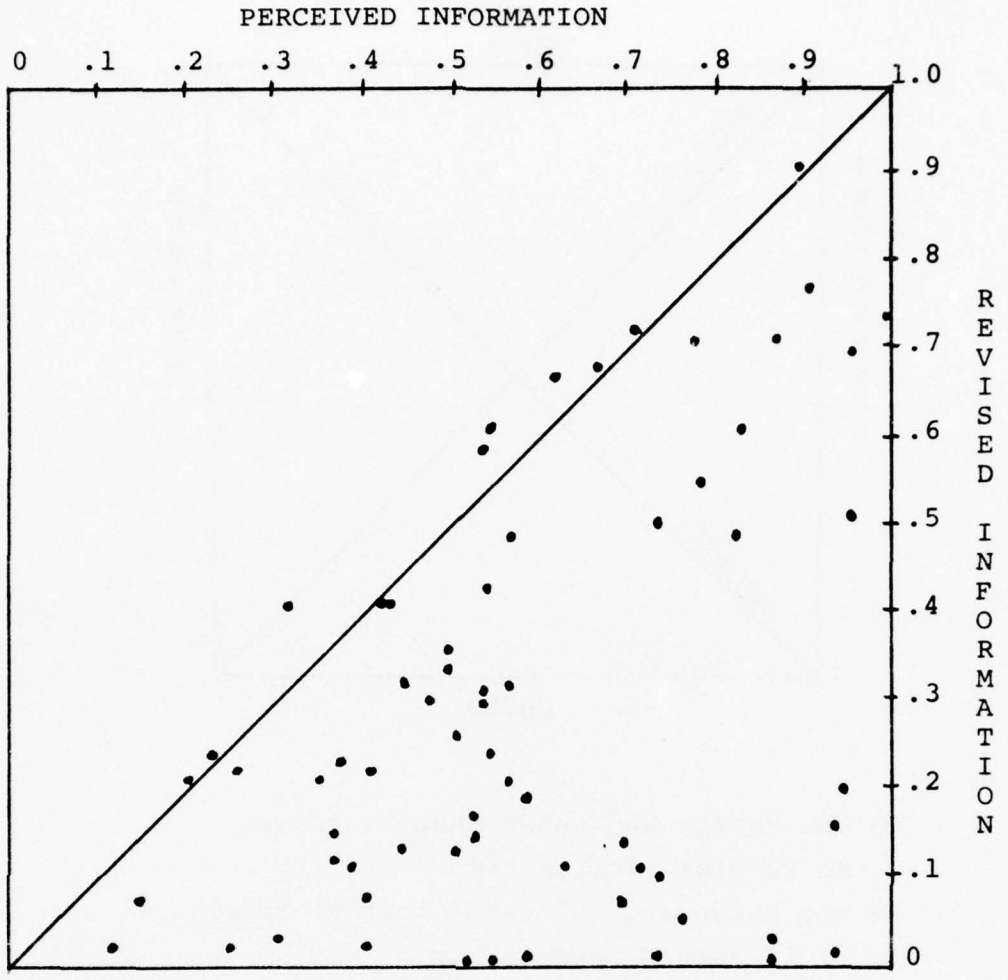


Figure 13. Relative Information on First Exposure (66 Subjects)

among their probability assignments. This seems to be a result of previous training in answering multiple choice items.

Figure 14 illustrates the amount of improvement in realism experienced by our subjects. Among the fifteen individuals who have taken a test on the computer two or more times, the average loss in score due to lack of realism on the first test was 108 points; the average loss in score due to lack of realism on the last test they took was 15 points.

THE FUTURE

The near term future of this work as we see it consists of at least three parts:

- o An on-going improvement of our test and interpretation ideas.
- o Several small-scale formal experiments to verify our design.
- o An implementation of our ideas on the University of Illinois' PLATO IV system.

This final point indicates our desire to make Admissible Probability Measurement available to a large community of potential users. Within that community we hope to encourage experimentation on a larger scale than we can support with our present facilities.

CONCLUDING REMARKS

At the beginning of the paper, I made reference to the possible importance of Admissible Probability Testing for selection and training in the military services. As a result of his experiment, Gardner [1] found the following

Tests	No. of Cases	No. Showing Improvement
1	8	N. A.
2	7	6
3	4	4
4	2	2
5	1	1
⋮		
11	1	1

Av. loss due to lack of realism on first test: 108

" " " " " " " " last " : 15

Figure 14. Effect of Experience on Realism Score

hypotheses to be supported:

- o The scores attained on a [probability] test will be a more accurate assessment of student knowledge than will choice scores.
- o [Probability] test results will be more reliable than choice test results since there is significant reduction in guessing.
- o [Probability] tests yield more information that can be used in policy-making for a multi-tracked curriculum than do choice tests.

We have taken these results as an encouragement to continue to develop and extend Computer Assisted Admissible Probability Testing.

ACKNOWLEDGEMENTS

I would like to thank Dr. Emir H. Shuford, Jr., and Dr. Thomas A. Brown, both of the Rand Corporation, for their obviously very important contributions to this work.

REFERENCES

1. Gardner, Willie C. (1969) The Use of Confidence Testing in the Academic Instructor Course. Proceedings of the 1969 Annual Meeting of the Military Testing Association, September 15-19, New York, N.Y.

2. Shuford, Emir H. (1969) Confidence Testing: A New Tool for Measurement. Proceedings of the 1969 Annual Meeting of the Military Testing Association, September 15-19, New York, N.Y.

3. Brown, Thomas A., Probabilistic Forecasts and Reproducing Scoring Systems, RM-6299-ARPA, June 1970, The Rand Corporation, Santa Monica, California.

CHAROSEL: A COMPUTER PROGRAM WHICH SELECTS QUALITATIVE PREDICTORS FOR QUALITATIVE CRITERION PREDICTION PROBLEMS

by

William J. Moonan, PhD

Director of Statistical Programs
Navy Personnel Research and Development Center (NPRDC)
San Diego, California 92152

A Presentation Made to

The 1973 Military Testing Association Conference
San Antonio, Texas

28 October - 2 November 1973

I. INTRODUCTION

A. Purpose

The author realized that statistical theory and practice had a serious gap; namely, that no adequate statistical technique existed for the analysis problem associated with the prediction of categorical membership from knowledge of other categorical (qualitative) information. He attempted (Moonan, 1972) to provide an adequate solution to this problem. His solution, augmented by a computer program, permitted the utilization of possibly several hundred available qualitative predictors. Since most operational situations employing prediction analysis cannot practically utilize predictor information of this magnitude, it was soon realized that some model seeking technique would have to be developed in order to make the ABCD technique more effective. Other prediction analysis procedures have their associative model seeking procedures as illustrated in Table 4, p. 17, of Moonan (1973). These other variable selection techniques were not useful for problems analyzed by the ABCD technique and, consequently, a new procedure called CHAROSEL was developed.

B. Background

The initiation of this development began with a real personnel problem, namely that of classifying individuals into various categories of drug abusers from knowledge of biographical, life-history, and opinion data items. It may be of interest to this

audience to note that both the ABCD and CHAROSEL procedures were conceived, analyzed, and programmed before data from the project were made available for analysis. There are several lessons to be learned from this experience, none of which can be elaborated here, other than the bemusing consideration that, contrary to many research analysts' opinion, personnel research planning and data collection activities are more time consuming than data analysis: Personnel research planning and data collection activities may indeed be much more time consuming than the development of data analysis techniques.

The type of prediction problem referred to has been labeled by the author as a Type D prediction problem. Other models for prediction analysis have been identified by Lubin (1950) and Rozeboom (1966) and are labeled as in Table 1.

Insert Table 1 about here

---

1. What is the kind of problem? Statistical consultants and analysts frequently find it useful to answer the question, "What kind of problem is this?" when consulting with clients or developing analysis procedures. The procedure has several advantages, since, if the question can be answered succinctly and clearly, the complications associated with the consultation or analysis are greatly simplified. Accordingly, we shall attempt to use this technique in our description. The answer to the question is most

appropriately the identification of a particular class of statistical classification techniques. Some of the candidates are indicated as follows:

a. The discrimination problem. Given a set of distinct populations and a multivariate random variable, the discrimination problem is that of defining a numerical index using the multivariate information in such a way that this index optimally discriminates among the populations. The classical solution to this problem is given by the linear discriminant function of R. A. Fisher (1936). The appropriate model seeking technique for this problem utilizes Mahalanobis' generalized distance statistic (1936) as illustrated by Rao (1952).

b. The taxonomy problem. If the researcher has reason to believe that his multivariate sample came from more than one distinct population, the taxonomy problem is used to evolve a rule for sorting the sample observations into several distinct populations. An approach to this problem solution for multivariate normal predictors has been provided by Wolfe (1967) and for simple binary attribute predictors by Tanimoto (1960). No specific model seeking procedures have been developed for these techniques or are known to the author.

c. The classification or diagnosis problem. Given a group of distinct multivariate populations for which samples are available, the solution to the classification problem is a decision rule that would enable one to assign an additional observation to his proper

population if the multivariate observational vector is available and it is known that the individual has been derived from one of the given populations. There are a variety of techniques available for application, all of which are based, in one way or another, on modern decision theory and, for the most part, utilize multivariate normal assumptions. Among these solutions are the Wald (1944) statistic and the Anderson (1951) approach. Model seeking approaches within the classification problem solutions are not generally available.

d. Types of variables. The concept of variables enter into the problem statement in two different ways. The first way occurs when consideration is given to the fact that we have both a criterion variable, i.e., category to be properly identified for an individual and a set of predictor variables from which the criterion category is to be ascertained. It is also interesting to note that our problem is uniquely concerned with variables that are described as "qualitative" rather than "quantitative." These types of variables differ with respect to the way by which their measurement is "manifested." Qualitative variable measurements are manifested by a specification of a quality of the variable on a nominal scale of measurement. When the variable is manifested by or for an object through an item of a questionnaire, the response quality is called an alternative. Quantitative variable measurement is manifested on an interval or ratio scalable, metrical dimension as a quantity. Our concern, in this report, is exclusively with qualitative variables used both as criteria and predictors.

From these considerations it definitely appears that the problem at hand is a classification problem whose solution, because of the qualitative nature of the predictor information, is provided by Moonan's ABCD technique for which, until now, no model seeking procedure was available.

2. Problem statement. Thus our problem, succinctly stated, is to devise a model selection technique for prediction problems involving the prediction (or decision) of the category of membership using as an information base for prediction a large set of other qualitative variables.

II. THEORY

A. Notation

θ_k represents the kth category of group membership, $k = 1(1)K$.

(1) $X_i = (x_{1i}, x_{2i} \ldots, x_{mi})$ represents the predictor qualitative variables associated with individual i.

(2) $P(X_i, \theta_k)$ represents the probability that individual i possesses predictor information X_i and is a member of criterion category k.

(3) $P(\theta_k)$ represents the a priori probability of a randomly selected individual being a member of category k.

(4) $P(x_m^a \mid \theta_k)$ represents a probability, known as an endorsement ratio, indicating the probability of an individual from category k responding to the ath alternative of the mth predictor character.

(5) Then $P(X_i \mid \theta_k) = \frac{P(X_i, \theta_k)}{P(\theta_k)}$ represents the conditional probability of $\underline{x}_i$ given θ_k. We are, however, primarily interested in the "inverse" of this probability which is obtained from Bayes' (1763) formula expressed as follows:

$$(6)\quad P(\theta_k \mid X_i) = \frac{P(\theta_k)P(X_i \mid \theta_k)}{\sum_{k=1}^{k} P(X_i)P(X_i \mid \theta_k)} \quad .$$

With this notation we are now in a position to describe the theoretical process of our model selecting procedure called CHAROSEL.

B. Procedure Description

1. General. The CHAROSEL procedure computes a vector of the Bayes' formula probabilities, $P(\theta_k \mid X_i)$, for each member of a testing sample for whom θ_k is known. This vector is vector multiplied by the column elements of a cost of misclassification matrix $\underline{C}_{\theta,\theta'}$. The elements of $\underline{C}_{\theta,\theta'}$ equal zero when $\theta = \theta'$. Otherwise, $\underline{C}_{\theta,\theta'} > 0$. This vector multiplication produces a set of $\underline{K}$ expected misclassification costs, $\underline{V}_k$, and the decision is that the $\underline{i}^{th}$ individual is a member of category θ_e if $\underline{V}_e$ is the minimum of the $\underline{V}_k$.

In order to maintain computational feasibility, it is assumed that the elements of $\underline{X}_i$ are stochastically independent so that:

$$(7)\quad P(X_i, \theta_k) = \prod_{m=1}^{M} P(x_{im}, \theta_k) \quad .$$

2. Variable selection technique. We shall accomplish the model seeking procedure by characterizing it as a problem of selection of the qualitative predictor variables. A similar approach to this class of problems has been frequently used by the author (Moonan, 1966, 1967, 1968). The optimal approach to this form of solution has long been considered by the author to be that of dynamic programming similar to the extremely clever solution to the clustering problem provided by Jensen (1968). Unfortunately, the dynamic programming solution to the model seeking problem has eluded me and I have had to be content, as in this case, with using what I call "conditional sequential optimization." This optimization procedure is defined as follows: The program selects predictors in such a manner as to optimize a meaningful objective function. The variable is first selected which, by itself, has the best objective function value; next, another variable is selected such that, when combined with the first selected variable, the two produce a predictor system with best conditional objective function value. Having found that pair, a new third predictor is sought which, when combined with the previous selected pair, produces a three variable predictor system with the best conditional objective function value. This process is repeated until all predictor variables are utilized.

3. CHAROSEL's objective function. In order to discuss the objective function of CHAROSEL, it is necessary to consider Table 2, which portrays the results from applying the ABCD technique for any

Insert Table 2 about here

given number of predictors in a three criterion category prediction problem. Table 2 may be interpreted as follows: There are three criterion categories labeled 1, 2, and 3, for which testing sample sizes of 39, 24, and 72, respectively, are available. Using ABCD for some specified set of predictor variables, 30 of category 1, 17 of category 2, and 61 of category 3 were correctly classified.

If the predictor variables were completely effective, the principle main - diagonal (30, 17, 61) would be represented by (39, 24, 72), showing that all testing sample members were correctly classified and all non-diagonal elements would have frequency = 0.

This suggests that the proper objective function is a statistic representing this condition. One such statistic is provided by an application of classical information theory and was suggested by the author to his programmer, Mr. Samuel Bowser. This statistic proved inappropriate for some applications and was modified by Mr. Bowser to equation (8).

(8) Objective Function

$$\frac{-\sum_{i=1}^{k}\left\{Pi\sum_{\substack{j=1\\ i\neq j}}^{k}P_{ij}\mathrm{Log}_2(P_{ij})\right\}+\sum_{i=1}^{k}\frac{n_{ii}}{f_{ii}}}{k}\quad \mathrm{Lim}\rightarrow 1.$$

$$\text{Lim} \quad f_{ii} \to 0 \frac{n_{ii}}{f_{ii}} = n_{ii}$$

P_i = A priori probability.

P_{ij} = Cell proportion off diagonal

n_{ii} = Number of occurrences in a given group

f_{ii} = Diagonal cell frequency

k = Number of groups

$\text{Lim } x \log x = 0$

$x \to 0$

Equation (8) has a value = 1.00000 if the decision tables' diagonal equals the frequencies of the testing samples' category sizes, otherwise the statistic is greater than unity.

III. IMPLEMENTATION

A. The CHAROSEL Technique: Computer Program Documentation

The CHAROSEL technique was programmed as a computer program for the NELC IBM 360/65 system and can be executed through the NPRDC IBM 1130 remote job entry system. The documentation for this program follows as extracted from Bowser (1973) and Moonan and Bowser (1973).

B. Identification

Title: CHAROSEL

Common Identification: CHARI

Category: Statistical

Programmer: Samuel E. Bowser

Analyst: William J. Moonan

Date: 1 May 1973

Compiler: IBM 360/65 FORTRAN G

Core Requirements: 375K

C. Usage

1. Input deck setup

a. Program deck and subroutine decks.

b. Input form including parameter and option cards.

c. Data either card or 9-track tape.

2. Input card layout

See input forms, Bowser (1973).

3. Equipment configuration

Input form cards are always loaded behind the program deck on the standard input unit. Input data can be on a separate tape or loaded from the standard input unit.

4. Output

a. Printer

(1) Decision table as each variable is selected.

(2) Values of objective function, change in objective function, total per cent correct decisions, and change in percentages of correct decisions.

(3) Variables are listed in rank order as they are selected.

b. Punch

(1) Endorsement ratios are punched out if they are calculated by program; this option is selected unless otherwise specified.

D. Requirements

1. Storage. The program as currently constructed does not utilize object time dimensioning. The size of the core requirement at present is 375K. This allows for 500 men and 200 variables. The number of categories is fixed at 10 or less and the levels of the variables is fixed at 10 or less. In order to exceed the size restrictions the redimensioning of the main program and its subroutines would be necessary.

2. Time. An approximation of the time required can be obtained from the following:

$$\text{Time (Min.)} = 5 + \left| \frac{(\text{Nr. of Variables})^2}{300} - 14 \right| (\text{Nr. of Categories} - 1)$$

Thus, for a three category, 50 predictor variable problem, the computing time of CHAROSEL for normal options is about 17 minutes.

For many other important details regarding documentation and instructions for filling out the input forms the user should consult Bowser (1973).

E. Program Restrictions

1. The training sample size must be less than or equal to 500.

2. The number of criterion categories must be less than or equal to 10.

3. The number of levels for each predictor character must be less than or equal to 10.

4. The number of predictors is limited by computer system core size and is related to sample size. Adjustment may be made to larger numbers of predictors but at present the program is set for 200 predictor variables maximum.

5. At present the program is operational only on an IBM 360/65 computer system and utilizes a core size of 300K bytes.

6. Prospective DOD users may arrange for program execution through:

Head, Computer Services Department

Navy Personnel Research and Development Center

San Diego, California 92152

7. Other potential users should contact the authors.

F. User Options

The user of CHAROSEL has considerable freedom with regard to type of procedure he wishes to utilize with this program. Among these options are:

1. The endorsement ratios may be either supplied by the user or the program will calculate them from the training sample.

2. A limit may be given for the maximum number of variables with missing data that are allowed for each object of the training sample

and/or missing data may be used as a "level" in the corresponding predictor variable.

3. Certain predictor variables may be forced to be utilized initially in the CHAROSEL predictor variable identification at the option of the user.

4. If the user wishes to utilize the cross validation process this may be accomplished by taking the variables selected by CHAROSEL and running them on the ABCD program with a testing sample.

5. Training sample data may be either on cards or on tape, however, there exists an input form to be completed by the user before execution.

G. An Illustration

1. The problem description. As an illustration of the CHAROSEL program, we present a problem actually executed on the program and provide the output results obtained thereby. This illustration is concerned with the educational problem of identifying intellectually gifted children who are members of so-called low socio-economic status groups within the city of San Diego. This illustration is the PhD thesis problem of Mrs. Charlotte Malone at the United States International University, San Diego. As yet, the thesis is incomplete and, of course, unpublished, and the reference to this work is only provided for illustrative purposes. The categories of criterion for this problem are two: θ_1, the children who have Stanford Binet IQ's greater than or equal to 132 and labeled "gifted" and, θ_2, those

children who have Stanford Binet IQ's less than 132 who are labeled "non-gifted."

The predictor information was derived from a questionnaire completed by the mothers of kindergarten children registering for school within a certain school district of San Diego during September 1973. The items of the questionnaire were composed of 51 biographical, behavioral, and life history items such as:

(1) Does your child use a lot of "grown-up" words? Does he talk like an older person than he really is?

A. Seldom

B. Occasionally

C. Often

D. Almost always

E. No answer possible

2. The results. The training sample for this problem is at present very small but will be augmented within a short time. For the present, N_1 = 28 gifted students and N_2 = 21 non-gifted students on whom data were available. The researcher specified the a priori probabilities to be $p(\theta_1)$ = .05 and $p(\theta_2)$ = .95. Each item of the questionnaire had five levels of response and their endorsement ratios, for each category, were computed by CHAROSEL. The execution of the program produced the following results: (See Table 3).

Insert Table 3 about here

We note that the decision table showed 17 misclassifications of gifted children and no misclassifications of non-gifted children and that the most optimal single predictor item was item number 1, shown previously. The objective function value was 1.78957 and per cent correct accuracy = (32/49)100 = 65.31.

The program proceeded to sequentially select items 29, 7, and 27, producing the information outlined in Table 4.

Insert Table 4 about here

At this stage, only eight misclassifications were made and the objective function has been reduced to 1.21067 and the per cent correct accuracy = $(\frac{41}{49})100 = 83.67$.

After 30 stages of selection the main - diagonal of the decision table was (28, 21), indicating that the per cent correct accuracy of classification was 100.00. The statistical results for this stage of analysis are available but not shown in this report.

The author has experienced the greatest personal satisfaction as a result of the utilization of his CHAROSEL technique to this identification problem which, if these preliminary results are sustained, will lead to better educational practices and improved well-being for many children.

The CHAROSEL program has been applied to several military and civilian personnel research problems. Some of these results are shown in Table 5.

Insert Table 5 about here

IV. CONCLUSION

The CHAROSEL technique represents a unique development in classification theory in that it possibly provides the only practical procedure available for selecting effective qualitative predictor variables for Type D prediction problems. A computer program, associated with this technique, has been developed and applied to several problems associated with both military and civilian personnel research. The results from these applications have shown the technique can identify a few predictor variables which will produce results of very high classification efficiency. This report provided information relative to the theory, implementation, and application of the program.

ACKNOWLEDGMENTS

The author would like to thank:

(1) Mr. Edmund Thomas, whose research on an NPRDC drug abuse project suggested the requirement for developing the ABCD technique.

(2) Dr. Norman Abrahams, whose work with program KEYCON and personal discussions suggested the need for the CHAROSEL program.

(3) Mr. Samuel E. Bowser for efficiently programming the CHAROSEL technique on the 360/65 computer.

(4) Mrs. Teddy Ralph, Mrs. Delores Callahan, and William K. Moonan for their assistance in writing this report.

(5) Mrs. Charlotte Malone for allowing me to discuss her PhD thesis problem and to present some data relating to it.

ADDENDUM

After this report was completed, the following thoughts occurred to the author.

The controlling aspect of CHAROSEL is its objective function. Suppose this was modified from equation (8) to the bivariate uncertainty measure

$$U(.,.) = - \sum_{i=1}^{k} \sum_{j=1}^{k} p(i,j) \log p(i,j) \ .$$

where p(i,j) is the joint probability of row i and column j in the CHAROSEL decision table. The objective of the procedure is, then, to select variables that maximize joint uncertainty, U(.,.). Used in this way the program could be used to select variables to affect heterogeneity among categories in the sense that the program will identify those variables which tend to make an independent relationship between predicted category and actual category.

This approach appears to have application to some analysis problems encountered at NPRDC.

REFERENCES

Anderson, T. W. (1951). Classification by multivariate analysis. Psychometrika, 16, 31-50.

Bayes, T. R. (1763). An essay towards solving a problem in the doctrine of chances. The Philosophical Transaction, Vol. LIII, 370-418.

Bowser, S. E. (1973). CHAROSEL I. Unpublished. Naval Personnel and Training Research Laboratory Statistical Analysis and Computer Applications Department.

Fisher, R. A. (1936). Use of multiple measurements on taxonomic problems. Annals of Engineers, Vol. VII, 179-188.

Jensen, R. E. (1968). A dynamic programming algorithm for cluster analysis. Management Science, 14, 1034-1057.

Lubin, A. (1950). Linear and non-linear discriminating functions. British Journal of Psychology, Section 3, 90-103.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. Proceedings of the National Institute of Science (India), Vol. 12, 49-55.

Moonan, W. J. (1972). ABCD: A Bayesian technique for making discriminations with qualitative variables. Proceedings of the 1972 Military Testing Association Conference.

Moonan, W. J. (1973). Vistas of analysis: Qualitative prediction and decision analysis. Proceedings of the 1973 Occupational Research Conference at the Navy Personnel Research and Development Center, San Diego.

Moonan, W. J., Balaban, J. G., Geyser, M. J. (1967). SEQUIN II: A computerized item selection and regression analysis procedure. U. S. Naval Personnel Research Activity. (Research Memorandum SRM 68-11)

Moonan, W. J., Beauregard, C. (1968). PARFORM: A computer program for developing statistically parallel tests of high validity. San Diego: U. S. Naval Personnel Research Activity. (Research Memorandum SRM 68-22)

Moonan, W. J., & Bowser, S. E. (1973). CHAROSEL Synopsis. Fact Document #27, Naval Personnel & Training Research Laboratory, San Diego.

Moonan, W. J., & Pooch, V. W. (1966). SEQUIN: A computerized item selection procedure. San Diego: U. S. Naval Personnel Research Activity. (Research Memorandum SRM 67-8)

Rao, C. R. (1952). Advanced statistical methods in biometric research. New York: John Wiley and Sons.

Rozeboom, J. H. (1966). The foundations of the theory of prediction. Homewood, Ill.: The Dorsey Press.

Tanimoto, T. T., & Rodgers, D. J. A computer program for classifying plants. Science, 1960, 132, 1115-1123.

Wald, A. (1944). On the statistical problem arising in the classification of an individual into one of two groups. Annals of Mathematical Statistics, 15, 145-162.

Wolfe, J. H.(1967). NORMIX: Computer program documentation. San Diego: U. S. Naval Personnel Research Activity. (Research Memorandum SRM 69-11)

TABLE 1

Types of Prediction Analysis

Independent Variables	Dependent Variable	
	Metrical Criteria	Categorical Criteria
Metrical Predictors	Regression Analysis Type A	Discriminant Analysis Type C
Categorical Predictors	Analysis of Variance Type B	ABCD Analysis Type D

TABLE 2

Illustrative Decision Table for a Three Category Problem

True Categories	Decision Category 1	2	3	Total	Percent
Category 1 Frequencies	30	4	5	39	28.9
Category 2 Frequencies	4	17	3	24	17.8
Category 3 Frequencies	6	5	61	72	53.3
Column Totals	40	26	69	135	
Column Percents	29.6	19.3	51.1		100

TABLE 3

CHAROSEL Decision Table for First Stage Analysis
For the Gifted Children Problem

True Categories	Decision Category 1	Decision Category 2	Total	Percent
Category 1 Frequencies	11	17	28	57
Category 2 Frequencies	0	21	21	42
Column Totals	11	38	49	
Column Percents	22	77		

Stage Selected	Variable Name	Nr.	Objective Value	Function Change	Percent Amount	Correct Change
1	01	1	1.78597	0.0	65.31	0.0

TABLE 4

CHAROSEL Decision Table for Fourth Stage Analysis for
The Gifted Children Problem

True Categories	Decision Category 1	2	Total	Percent
Category 1 Frequencies	20	8	28	57
Category 2 Frequencies	0	21	21	42
Column Totals	20	29	49	
Column Percents	40	59		

Stage Selected	Variable Name	Nr.	Objective Value	Function Change	Percent Amount	Correct Change
1	01	1	1.78597	0.0	65.31	0.0
2	29	29	1.38743	0.39855	75.51	10.20
3	7	7	1.24807	0.13936	81.63	6.12
4	27	27	1.21067	0.03740	83.67	2.04

TABLE 5

Recent NPRDC CHAROSEL Applications

Problem Name	Criterion Categories	Types of Predictors	Training Sample	Nr. of Predictors Used	Objective Function Value	% Classification Accuracy
Drug Abuse Prediction	(3) A. Hard Drug User B. Recreational Drug User C. Abstainers	(139) A. Biographical B. Opinion Data	(362) Naval Recruits	(5) (18)	1.90207 1.72074	61.6 68.2
Odds for Effectiveness	(2) A. Recommended for Reenlistment B. Not Recommended	(140) A. Biographical B. Self-evaluation C. Opinion	(320) Naval Recruits	(7) (19) (95)	1.38476 1.11748 1.00000	95.6 98.4 100.0
Skid Row Employment	(2) A. Completed Project B. Dropped Out	(52) A. Biographical B. Self-evaluation	(179) Skid Row Welfare Recipients	(13) (39)	1.39013 1.26212	85.5 88.8
Target Classification	(3) A. U. S. Subs B. Foreign Subs C. Surface Ships	(19) Sonar Clues	(91) Sonar Contacts	(1) (4) (19)	1.51500 1.07334 1.07334	70.3 95.6 95.6
Identification of Gifted Children	(2) A. Gifted Children B. Non-Gifted Children	(51) Behavioral Characteristics Questionnaire	(49) School Children in Kindergarten	(1) (4) (30)	1.78957 1.21067 1.00000	65.31 83.67 100.00

A FACTOR ANALYTIC APPROACH TO THE CRITERION PROBLEM

ROGER PENNELL

Air Force Human Resources Laboratory
Lowry AFB, CO 80230

November 1, 1973

A FACTOR ANALYTIC APPROACH TO THE CRITERION PROBLEM

I. Introduction

The problem of how to deal with multiple criteria when attempting test, or test battery, validation is a topic of continuing concern (e.g. Blum and Naylor, 1968; James, 1973; Nunnally, 1967; Roseboom, 1966; Schmidt and Kaplan, 1971). In the most complex case, the predictor set (PS) and the criterion set (CS) contain multiple measures. The development of an optimal set of weights to obtain a maximal correlation between the two sets is well known from the theory of canonical correlation (e.g. Morrison, 1967). This solution, however, answers the wrong question. In obtaining measurement on a CS, one attempts to utilize variables which measure a construct, namely, "goodness" of job performance, say. Thus, the intent is to find a weighting of the CS which maximizes the congruence of the criterion composite (CC) with this hypothetical construct. One begs the question if he employs a canonical correlation approach, since consideration of the job performance construct is lacking.

A paper by James (1973) takes construct validity as a point of departure in discussing the criterion problem. He considers four models: the so-called ultimate criterion model which combines all available criterion measures into a single composite, the multiple criterion model which seeks to better understand the interactions between the PS and CS by maintaining the individuality of the CS measures, the general criterion model which merely formalizes the notion that the criterion construct is likely to be complex and multidimensional, and the construct validation model which

emphasizes that the criterion problem should really be couched in a construct validation framework. James rejects the ultimate criterion model as being "erroneous", perhaps making the tacit assumption that any such criterion cannot have the requsite construct validity. Furthermore, he suggests that "advanced" methods of analysis, such as discriminate analysis, obviate the necessity of ultimate criterion development and, thus, one is naturally led to a multiple/construct valid kind of approach. It is completely unclear to this author how discriminate analysis is a useful tool for either construct validation or PS validation. It is precisely this kind of superficial analysis which provides the fuel necessary to sustain a potentially reconcilable controversy. Construct validity, indeed, is a totally separable issue apart from the ultimate/multiple criterion issue. The fact is that construct validity is a premise, not an arguable point (how to achieve it is another matter). With this premise in mind, a reasonable and desirable goal of a practitioner would be to explore CS predictability; certainly, discriminate analysis is not the answer unless there is but a single criterion measure which can be dicotomized on a success-failure dimension. Furthermore, it would seem that ultimate criterion advocates are regarded as carefully avoiding the CS data until an assistant has combined it into the required form, thus obscuring or eliminating potentially interesting relationships. Certainly, the practitioner will explore the relationships existing in his data to as detailed a level as his skill or computing capability permits, whether or not he finally elects to form a single CC.

This paper assumes the construction of a so-called ultimate criterion is the goal, and will provide a method of accomplishing this which is, in some sense, optimal. The assumptions concerning construct validity will be delineated by way of a two dimensional framework for the CS which provides natural implications for the appropriateness or inappropriateness of various prediction considerations.

2. Conceptual Framework for Criterion Variables

The first dimension of interest for conceptualizing the CS concerns the number of measures in the set. In the simplest case the CS is but a single measure; subsequent analysis, obviously, will depend solely on the adequacy of this measure. Validation efforts using the PS utilize either bivariate linear regression, if N(PS)= 1,* or multiple linear regression, if N(PS)>1. This situation is obviously the most desirable from a computational point of view, however, the results may be totally irrelevant, if the criterion measure lacks construct validity. The other end of this dimension is defined by the case N(CS)>1. Here the classical problem of weighting the CS to obtain the "best" criterion measure is encountered.

The second dimension important in conceptualizing the CS, which was alluded to above, is the extent to which the CS can be regarded as having construct validity. Here we are concerned with the congruence of the CS and the construct of interest. If N(CS)= 1, and this single criterion measure can be regarded, either conceptually or experimentally, as having construct validity, we have obtained the best of all worlds: validity coefficients obtained using the PS demonstrate relationships with the actual construct of interest, and problems of CS

---

* The notation N(X) = Y shall be read as the number of elements in the set X = Y.

weighting are avoided. When N(CS)>1, one of three situations can obtain: Case I, the CS can be regarded as an exhaustive set of measures of the construct. Here one must judge that no additional measure could add predicitive validity relative to the construct. Case II, the CS may be regarded as a random sample of measures potentially related to the construct. This is a situation similar to test construction where an attempt is made to include items which constitute a more or less random sample from the domain of interest. Case III, the CS may be regarded as an at-hand or non-random sample of measures potentially related to the construct.

This two dimensional structure for the CS makes it clear where potential solutions for criteria weighting are possible and useful. When N(CS)= 1, the only concern surrounds construct validity. When N(CS)>1, solutions are possible only for Cases I and II. For Case III, no weighting, optimal or non-optimal can produce a CC suitable for PS validation.

The following section will briefly outline the nature of solutions which have been proposed for N(CS)>1, mostly without regard to whether Case I, II or III is appropriate.

3. Historical Solutions for N(CS)>1

The first class of solutions will be called Simple Weighting Solutions. The least elegant member of this class is the CC formed by using equal weights for the CS set. This solution attempts to optimize no thing, neither validity with the PS nor construct validity. A common approach, similar to the one above, is to use nominal weights developed by subjectively evaluating the probable importance of each

CS measure relative to the construct of interest. This solution attempts to maximize the extent to which the CC measures the construct, thus, optimizing construct validity. Another member of the simple weighting class uses the reliability of each CS member as its weight. This method takes advantage of the fact that reliability is directly a function of true score variance, i.e. the higher the reliability, the higher the relative true score variance. Thus, such a weighting scheme maximizes the proportion of true variance, under the constraint that all the CS measures are included in the CC.

All of the simple weighting methods suffer at a minimum from a lack of consideration of the matrix of correlations among the CS. If the CS exhibits any appreciable cross-correlations (as they in general will if construct validity has any tenability), then any simple weighting of the individual variables will produce a CC of unknown composition.

Often, any system of weighting which permits a different configuration of scores to achieve the same CC score is not justifiable. An example would be a situation in which a minimum performance on each member of the CS must be obtained before performance is considered acceptable. This would be the case for pilot performance, for instance, where minimum levels of accomplishment for all performance components must be attained. The method commonly employed is the multiple cut-off method. Each component of the CS is usually judged to have obvious content and construct validity.

Other approaches, here called empirical weighting, use weights which maximize a mathematical criterion. An example,

previously alluded to, is the method of canonical correlation. This method suffers the fundamental flaw of ignoring the construct while maximizing predictability. An empirical weighting scheme developed by Kurtz (1937), for the Case N(PS)>1 and N(CS)>1, develops a set of weights for the PS such that the so weighted composite PS is maximally related to each member of the CS. This method seems to answer a question one would tend not to ask, since we discover neither the maximal relation of predictors so weighted to any of the criterian measures, nor which of the PS are actually functioning as predictors. A related method, which has considerable theoretical justification in practice, is the utilization of weights which are proportional to the first principal component of the CS. The first principal component of a test battery is produced by finding a set of weights (with unit variance) such that when applied to the origional measures, we produce a new variate having maximal variance. For Case I and II, this approach defines a theoretically defensible method for constructing a CC, providing the criterion domain is unidimensional. The latter consideration is no doubt rarely the case in practice, i.e. most would agree that job success, say, represents a multidimensional system such that many facets contribute to success (cf., Ghiselli, 1964). These considerations lead us to the question of the factorial complexity, in a factor analytic sense, of the criterion, and to a general solution of the criterion weighting problem.

4. Factor Analytic Weights for Multiple Criteria

The theory of psychometric inference will be helpful to our development of a general weighting scheme. Psychometric inference is

more readily understood when contrasted with classic statistical inference. Statistical inference attempts to make judgements about the value of a parameter, or set of parameters, in a population of potentially measurable objects (people, light bulbs, plots of land, etc). These inferences are made by selecting a random sample of these objects, and using the value of the parameter in the sample to make inferences about the parameter in the population. Psychometric inference, on the other hand, seeks to make inferences about a population of tests or measures based on a sample of tests at hand (cf., Kaiser and Caffrey, 1965). Specifically, we wish to know if results from a factor analytic study, i.e. a set of factors, has generalizability in the sense that another sample of tests from the same domain or population would tend to render the same factors. Thus, we make inferences about population factors from a sample of tests defining the factors in the population.

We can be quite sure about the generalizability of our results if we sample exhaustively from the population, or, obtain a sufficiently large random sample of measures. These are exactly the conditions required for Case I and II construct validity. The question can then be asked, suppose we obtain a factor analysis of the CS, assuming Case I or II holds. Is there a theoretically justifible CC which can be constructed from this factor analysis? The answer is yes, and the CC is constructed as follows (for equational development see Appendix I).

Factor analyze the CS (some methods of accomplishing this are preferable to others); rotate the obtained solution to a psychologically

pleasing orientation (this can be accomplished using one or another analytic procedures such as Varimax; Kaiser, 1958). The next step is to compute the associated factor scores based on the rotated factors. These scores represent scores for people, on measures (factors) derived from linear combinations of the basic, observed test scores. Now we estimate factor saliencies by computing the sum of squared factor loading for each factor. These saliencies will represent a set of weights for the factor scores. Let the CC be defined as the sum of the weighted factor scores.

The model we are proposing says that when case I or II can be assumed to hold, the results of a factor analysis of the CS can be said to estimate the test domain factors. Given this, the weighting problem is solved as a natural byproduct of the analysis, namely, the weights are defined by the estimated importance of the factors in the domain. It should be noted that these weights are *relative, and not invariant or absolute.* The magnitude of the weights will be governed by the initial scaling imposed on the covariance matrix of the CS. If the CS correlation matrix is factor analyzed, the weights will be different in magnitude from the factor weights from the covariance solution, say. However, one typically does not ask for absolute invariance, but rather, invariance up to a multiplicative scaling.

Note that the present model includes as a special case the first principle component weighting scheme. When the criterion domain is found to be unidimensional, or to contain one common factor, weighting the tests according to the elements of this component is identical to producing a set of factor scores based on one common factor.

It should be noted, as well, that a common factor analysis is implyed by the present model. Common factor analysis attempts to analyze variance common to two or more variables. This is accomplished by estimating the unique variance contained in each measure in the CS, and partialing this variance out before the analysis is begun. (cf., Harman, 1967; Muliak, 1972.)

5. Discussion

At this point it seems prudent to clear up a confusion which occasionally occurs in the literature (cf. Schmidt and Kaplan, 1971). The confusion surrounds the distinction between factorially pure measures and orthogonal or uncorrelated measures. The Schmidt and Kaplan article attempts to argue along the lines that it is generally conceded that simply weighted composites result in lower validity unless the measures are factorially pure. But, they continue, the validity of a test (defined as the sum of the cross products of the appropriate factor loadings) can under various circumstances be higher for factorially complex tests. Therefore, the composite will not in general have lower validity.

The confusion evident in this reasoning is difficult to sort out entirely, but seems to be based on the distinction between factor pure measures and uncorrelated measures. A measure can be factor pure and still be correlated with another variable, similarily factor pure, because they load on the same factor. In any event, it is safe to say that a composite variable, formed from correlated or uncorrelated component variables, using any simple weighting rule, will have decreased validity from the theoretically maximum. The composite having maximal validity

with the PS is defined by the weights estimated via canonical correlation, if N(PS)>1, or the multiple regression of the CS on the predictor, if N(PS) = 1. This is plainly obvious when N(PS) = 1; any set of weights departing from regression weights for the CS produces lower validity, since these weights maximize the obtained correlation. However, these considerations are largely misplaced since one is not interested in optimally weighting the CS to produce maximal validity with the PS. What we are attempting is to optimally weight the CS so as to be maximally congruent with the relevant hypothetical construct. It seems to be a case of misplaced validity.

Related to this matter is the issue of the so called criterion oriented approach to test construction (Nunnally, 1967). Here, one attempts to construct items for use in a predictor test which are maximally correlated with a known criterion. Such a set of items can then be expected to produce maximal validity between a predictor, constituting these items, and the criterion. Nunnally (1967) argues that this criterion oriented predictor is likely to be very complex factorially, and lead to little conceptual understanding about the attributes which are important for criterion prediction. He argues that off-the-shelf, pure measures are more likely to increase our understanding of the entire process, as well as producing comparable validity. This kind of thinking apparently prompts Schmidt and Kaplan (1971) to recommend that increasing the factorial purity of both the PS and CS will tend to resolve the criterion problem in general.

This latter argument makes the tacit assumption that decreased factorial complexity on both sides of the equation will result in decreased complexity in prediction. Indeed, the independent construction of pure measures on both sides, will, in general, not result in simpler predictive relations unless the factors on one side happen to be isomorphic to those on the other. This may occur in practice, but surely constitutes an assumption to be verified. Furthermore, striving for factorially pure criterion measures may be a laudable goal, but it is often the fact that criteria constitute givens. Just satisfying the conditions for Case I and II may tax the imagination of an innovative test designer without striving for the nicety of factorial purity as well.

Achieving clarity in the multivariate relationship between the PS and the CS constitutes the primary motive underlying theoritical discussions about the factorial composition of the two data sets. Considerable insight along these lines can be attained using the method of factor extension (Dwyer, 1937). Once a satisfactory factor solution has been obtained for the CS, preparatory to constructing the composite score, factor extension techniques can be used to show how the PS measures relate to the CS factors (see Appendix II). This method uses the cross correlations between the PS and CS, in conjunction with the CS factor loadings, to compute the loadings of the PS on the CS factors. The end result is a factor matrix for the PS where the factors are defined by the CS analysis.

This procedure permits definition of the PS variables in terms of broad, uncorrelated factors, and should facilitate theoretical development by permiting unambiguous coupling of the PS and a set of theoretical

constructs (factors). It should be noted that the factorially complexity of neither the PS nor the CS is of paramount importance. So long as the factor solution of the CS is interpretable psychologically, the procedure is workable and provides substantive information about the PS-CS relationship.

6. Summary

This paper has reviewed the various methods of criterion construction under conditions of single and multiple criteria, as well as the theoretical justification for these procedures, and has developed a two dimensional framework for conceptualizing the criterion problem. It was noted that controversies surrounding, (1) construct validity, and (2) the multiple vs single criterion camps seem to be of a "straw-man" nature. Based on the review and the development of the theoretical framework, a solution for the composite construction problem was proposed which may have optimal properties. Certain implications of the model in terms of interpretation and factorial complexity of the PS and CS measures was discussed.

TECHNICAL APPENDIX I

For the following development define

N: the number of observations; sample size for the CS and PS

n. the number of variables in the CS

p: the number of factors

q: the number of variables in the PS

Let R be the nxn, correlation matrix among the CS variables. Compute A, nxp, which is the common factor matrix for p factors based on the canonical factor analysis algorythm (Rao, 1955), such that

(1) $AA' = U^{-1}RU^{-1}$

holds in an approximate sense. U^2 is the diagonal matrix of unique variance estimates. This is the usual factor analysis model, postulating a rank p factor matrix to approximate the common portion of the data. The canonical algorythm is chosen because of the desirable form of the resulting factor scores (Harris, 1968). However, other methods, such as principal factor analysis, could be used with little impact on the resulting CC. At this stage we generally require a transformation on A, say T, pxp, which provides a more easily interpretable solution. Thus, let

(2) $A^* = AT$

and

(3) $F = ZR^{-1}A^*$.

F, Nxp, is a set of factor scores estimated using the so called regression method (Harmon, 1967), where Z, Nxn, is the original scaled score matrix such that

(4) $Z'Z = R$.

Now, if we let $f_{ij}, i=1,2,\ldots,N, j=1,2,\ldots,p$, be the factor score for the i-th person on the j-th factor, the CC variable, C_i, is defined as

$$(5) \quad C_i = \sum_{J=1}^{p} W_j f_{ij},$$

where W_j represents the sum of squared factor loadings for the j-th factor of A*. It should be noted that F constitutes a set of scores which are uncorrelated, i.t.

$$(6) \quad F'F = D$$

where D is a diagonal matrix. In practice F would be normalized to unit length before (5).

TECHNICAL APPENDIX II

Let R_{pc}, q x h, be the cross correlations between the PS and the CS. We wish to find a matrix A_p such that

(1) $A_p A^{*\prime} = Rpc$,

i.e., A_p is a set of factor loadings for the PS variables on the CS factors. The solution is (Dwyer, 1937),

(2) $A_p = Rpc\ A^*(A^{*\prime}A^*)^{-1}$.

REFERENCES

Blum, M. L. and Naylor, J. C. Industrial Psychology.
New York: Harper and Row, 1968.

Dwyer, P.S. The determination of the factor loadings of a given test from the known factor loadings of other tests. Psychometrika,1937, 2, 173-178.

Ghiselli, E. E. Dimensional problems of criteria. In Readings in Psychological Tests, Barnette Jr, W. L. (Ed).
Homewood, IL: Dorsey, 1964.

Harman, H. H. Modern Factor Analysis.
Chicago: University of Chicago Press, 1967.

Harris, C. W. On factors and factor scores.
Psychometrika, 1967, 32, 363-380.

James, L. R. Criterion Models and Construct Validity for Criteria.
Psychological Bulletin , 1973, 80, 75-83.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis.
Psychometrika, 1958, 23,187-200.

Kaiser, H. F. and Caffrey, J. Alpha factor analysis.
Psychometrika, 1965, 30, 1-14.

Kurtz, A. B. The simultaneous prediction of any number of criteria by a unique set of weights. Psychometrika, 1937, 2, 95-101.

Morrison, D. F. Multivariate statistical methods
New York: McGraw-Hill, 1967.

Mulaik, S. A. The foundations of factor analysis.
New York: McGraw-Hill, 1972.

Nunnally, J. C. Psychometric theory.
New York: McGraw-Hill, 1967.

Rao, C.R. Estimation and tests of significance in factor analysis.
Psychometrika, 1955, 20, 93-111.

Roseboom, W. W. Foundations of the theory of measurement.
Homewood, IL: Dorsey, 1966.

APPLICATIONS OF PREDICTOR ORDERING AND SELECTION BY A BAYESIAN-DECISION TECHNIQUE

S. E. Bowser

Navy Personnel Research and Development Center

San Diego, California

Introduction

The selection of predictor variables has been the pastime of researchers since the inception of behavioral science research. We have tried to find the variables which are the "valid" predictors of criteria for our problems. In a prediction problem there usually exists a need to use some character or variable selection process to reduce the number of variables required to make predictions in operational situations. The concept of selection engenders the idea of ordering the variables so that the "best" subset can be selected for operational utilization. The problems associated with predictor ordering and/or selection are multidimensional.

The nature of the ordering and selection methodology used is dependent upon the types of variables being considered. The analyses areas can be characterized by the nature of the criteria and predictor variables. The

---

Insert Table I

---

type A predictor ordering and selection problem is generally handled by accretion-deletion techniques (Johnson, 1949, p. 331) within the general class of regression problems. Multiple regression is commonly used in personnel research, often by making unwarrented assumptions about the variables being used. The reason behind this tendency to over use regression methodology

lies in the fact that the technique is readily available in most modern computer centers and other techniques are not as available.

The type B ordering and selection problem is normally dictated by theory and/or practical considerations. It so happens that the analysis of variance models can be related to the analysis of data collected by design of experiments techniques and also can be used for type B prediction. The analysis of variance models are related directly to the multiple regression models. The principal difference lies in the independent variables being quantitative in type A prediction and qualitative in type B prediction. The distinction of quantitative versus qualitative is often within the control of the experimentor due to the general availability of scaling methodology.

The type C ordering and selection is defined only in terms of theory and logic, not methodology. Prediction analysis is possible through Linear Discriminant Function (Fisher, 1936). In specific cases the generalized distance function D^2 of Mahalanobis is the best methodology. In general this type of problem is converted by assumption to type A and the ordering and selection done via that methodology.

The type D type problem is the area in which there are the fewest number of methods generally available. In terms of prediction the area of information theory gives general guide lines but there is no specific methodology. Order statistics provide methodology for prediction analysis but are of little help in the ordering and selection problem. A new approach via Bayesian-Decision theory has been established. The computer program for this methodology is titled CHAROSEL (Bowser, 1973a). The CHAROSEL technique developed by Bowser and Moonan (1973) is based upon the Attribute Bayesian Classification Decision (ABCD) technique developed by Moonan and presented

in its initial form to this conference last year.

The application of the ABCD technique to type D problems is as follows. The probability of the observed category of predictor variable arising as a result of sampling each criterion category is first calculated. By application of Bayes' formula, this gives rise to posterior probabilities for each subject of the testing sample. These probabilities are then combined with a cost matrix leading to specific probability of criterion group membership. Whichever decision minimizes costs of misclassification becomes the predictive class for the sample objects. Such a strategy of decision is known as a Bayesian-Decision technique. CHAROSEL is the technique for selection and ordering of categorical predictor variables for categorical criterion prediction problems.

The application of CHAROSEL to type D selection and ordering problems will be presented in some detail. The population of interest is sampled in order to obtain a training sample. The purpose of this sample is to estimate the probabilities of endorsement of each level of each predictor variable for every criterion category. The prediction is effected by using a cost of misclassification matrix and posterior probabilities of criterion membership estimated by endorsement ratios and all combined via Bayes' theorem. The resulting decision table is determined by Bayes' theorem probability, minimum expected cost of misclassification, and minimum objective function.

These decisions are then assembled into a decision table whose rows represent known categories of membership and whose columns represent predicted categories of membership for members of the testing subsample. The training and testing subsamples are expected to be independent for cross validation purposes and also they should be large enough to accurately estimate endorsement probabilities. The CHAROSEL program produces a new decision table for each predictor variable added to the prediction system. The program

selects the "best" variable at the time but it is an accretion system only in its present form. The nature of the objective function used has been discussed earlier by William Moonan in his presentation.

The development of the CHAROSEL program is a breakthrough in that predictions of qualitative criterion variables from knowledge of qualitative predictor variables is not possible. The ordering problem is also answered by application of this technique. This procedure allows the utilization of variables with no assumptions being made about their distribution or scaling properties. The only assumption that is made is that the predictor variables are independent which is an assumption that is made by most theories when this type of material is analyzed.

Application

The areas for which the CHAROSEL procedure has had application are both diverse and interesting. This methodology can be applied to any situation in which an attempt is made to predict categorical criteria from categorical predictors. In most cases the applications have dealt with non-cognitive data bases. It is interesting to note that the technique is not limited to behavioral science applications, but has had impact on other areas of research. The areas of potential application are expanded if one just observes that many continuous variables are naturally categorical variables which have been scaled.

Insert Table II

An overview of specific applications are listed in table II. The first two items listed in the table are derived from the project Classifi-

cation Prior to Enlistment. The findings reported represent initial findings rather than cross-validated results. The final results of this project have recently become available and show encouraging results in the cross-validation (Bowser, 1973b). The principal point here is that the criteria are of great interest to the military researcher and the data source is relatively untapped. The non-cognitive data bases represent a broad area of reliable predictor variables which are easy to obtain.

The next item on the table is a non-military project conducted in a welfare agency on skid row in Los Angeles (Bowser, Sherman, & Whisler, 1974). The results reported in the table are an attempt to look at whether or not non-cognitive factors could be used for screening successful from unsuccessful individuals in this type of project. It is apparent that biographical data has potential for this type of population as a source of prediction variables (Bowser, in preparation).

The target classification research is the next item on the list. This project understandably has great importance to the Navy. The details of the variables used are not available but the excellent results can be pointed out. This project demonstrates the versatility of application of this methodology. It is a good example of a non-behavioral science data base to which this procedure has been successfully applied. It is also obviously an excellent example of the advantage of being able to utilize cost of misclassification. In the target classification the cost of misidentification of foreign submarines as friendly is much greater than the cost of classifying our submarines as surface ships. It is thus restated that the potential of CHAROSEL to directly use a cost matrix is of a definite advantage.

The next item on our list is a good example of the type of application which usually would have been pushed into the regression mold. The flexibility questionnaire is the type of instrument that is usually scaled into a continuous variable. The other approach is to break it down into zero-one data and apply regression analysis. The results reported in the table show that when utilizing CHAROSEL the prediction of the criteria is as good or better than the expected results from the more traditional analysis procedures.

Considering page two of table II, the first item is interesting in that the criteria is polychotamous, a difficult situation for linear regression to handle. The point is that the CHAROSEL technique handles this problem with relative ease. The results reported here are in this case preliminary, as this project is continuing at the present time. This limited application was utilized to verify some theoretical ideas as to the questions which were important to this prediction problem; a definately useful approach for this technique.

The application of this methodology to a standard test such as the Strong Vocational Interest Blank is the next item on our list. The use of new methods on old problems is, or at least can be, a good thing. It sometimes allows a new viewpoint to be established. In the case reported here the results are encouraging. Please note the maximum correct decision rate which is 95.4%. This result is not cross-validated and so should be viewed with some caution but it is still very good and demonstrates a real promise for applying this method to established test instruments.

In the next application, the criteria was artificially categorized. Since the usual case is in the opposite direction this is somewhat interesting. The results as obtained are quite gratifying. The particular problem of reserve retention addressed here is one of great interest to all military

researchers. Again this method may allow the use of data sources previously ignored or not recognized as demonstrated in this case.

The identification of gifted children is the last project on our list. It gives us a clue to potential test construction approaches that are less culturally biased than our present methodologies. It should be noted that this study was conducted with a population consisting of children principally from minority groups. This study was reported in more detail by William Moonan in this proceeding.

The last item on this table demonstrates an approach which obviously has great potential value. Of course the details of the predictor variables are unavailable but the results are impressive. If nothing else, the implementation of these results will reduce the amount of time needed to board commercial airliners, a problem I'm sure is of interest to all those here.

As has been stated the actual applications have covered a broad range. The obvious versatility of this methodology demonstrates its great potential to researchers. The applications to date have included military as well as civilian research problems. The problem areas have been both in and out of the behavioral sciences. The data bases utilized have included sonar clues, biographical inventories, and vocational interest inventories.

Potential Applications

It is fairly obvious from the general applications already utilized that new applications are almost unlimited. Areas of application that readily come to mind are those of the medical field. The potential of this technique as a diagnostic tool is apparent. With further thought we can also see the possibilities of utilizing this method for diagnostic syndrome detection.

The possible applications to this area could take many forms, such as the research ideas just stated, or in the more direct application of providing a physician with probabilities of diagnostic category. Since the medical profession is currently using physician's assistants to help reduce the work load for the doctors, any technology which helps to objectify diagnostic procedures has great value.

The military applications already reported include the area of recruit screening and selection. An expansion in this area of application could possibly be made in the area of prediction of career field success. Since the studies in this area have utilized biographical information, it is also possible as a byproduct of this research to develop a more complete biographical profile of the successful military recruit. Investigation of such a profile would allow the military services to gain a better understanding of the types of individuals who are successful in military service. The understanding of the individual types which are successful would allow more reasonable manpower planning and provide the basis for modeling proposed alterations in screening techniques.

The area of organizational development and structure could also be affected by application of this technology. While it is not obvious, it is possible to use this methodology to uncover managerial value systems. Another related area lies within the realm of implicit judgemental criteria, for instance, personnel rating systems. It is further possible, by utilizing personal probability estimates, to use this method in obtaining data concerning decision processes themselves. This of course has great importance to managerial systems. Perhaps another area which is related to the foregoing should be explored, that is, the determination of implicit versus explicit goals of an organization.

While this technology is primarily concerned with categorical information as source material, it is recognized that mixtures of continuous and categorical information are likely. One possibility for handling this situation using the present computer program would be simply to categorize the continuous variables. This procedure, while the reverse of what is usually done, has certain features of arbitrariness which make it undesirable. It is possible to obtain probabilities from individual continuous scores. The method utilizes the mathematical features of the density function. There has been a program completed called MIX, (Moonan & Covher, 1972) which was developed by William Moonan; this program does combine categorical and continuous variables in order to obtain a decision table. It should be noted that this procedure is not comparable to CHAROSEL since it does not have the selection and ordering feature. It is possible to adapt CHAROSEL to use both continuous and categorical predictor variables. The proposed new program, CHAROSEL II, would broaden the applicability of this technique considerably. Other alterations of this approach are potentially possible, such as altering the objective function in order to obtain items which were the least discriminatory. The finding of items with great communality in some circumstances is more important than finding those items which discriminate. It is possible to weight the objective function so that a given class of prediction is made more important. This is over and above the use of a cost of misclassification matrix.

The potential applications of the present system are extensive. The system obviously can be expanded. The possibility is there to broaden the applications of this technique into the interface area between type C and type D prediction problems. The actual expansion of this technique into its many potential applications will of course depend upon the information of its availability reaching the appropriate researchers.

CONCLUSIONS

This technique is indeed a breakthrough in the area of predictor ordering and selection techniques. It provides ready access for the researcher into vast new data sources previously unavailable. The fact of the matter is that with the present anti-test pressures that are being exerted, new data sources are virtually imperative. The technique as described offers a technology capable of utilizing the non-cognitive data bases.

The method, while offering promise, still has drawbacks. It is not readily usable without the aid of large-scale computers. Operational systems based upon this technique would require computer support. Another problem, not quite so obvious, is the fact that this system lacks intuitive appeal. The logic of the procedures used is not readily obvious to individuals who are not familiar with Bayesian inferential systems. Given these drawbacks and possibly others not yet recognized, this method is still a promising new approach.

It is hoped that the introduction of this new approach will lead to broad new applications. The technique itself obviously needs to be explored in more depth. It can be shown that this technique is the same as linear regression if the assumptions are met. However, since no linear assumptions are necessary for this approach it is assumed that it is applicable where the linear model is inappropriate. It is intuitively obvious that nature does not always proscribe the linear model. It is hoped that this procedure will be explored more completely for its theoretical properties. Meanwhile perhaps the application of CHAROSEL will give added dimension to the formation of predictive systems and theories.

References

Bowser, S. E. (1973a) CHAROSEL I. Navy Personnel and Training Research Laboratory: Computer program documentation.

Bowser, S. E. (1973b) Non-Cognitive Factors as Predictors of Individual Suitability for Service in the U.S. Navy; Technical Report in press Navy Personnel Research and Development Center, San Diego.

Bowser, S. E.; Sherman, G.; and Whisler, R. H. (1974) An Action-Research Approach to Central City Unemployment. Journal of Vocational Behavior, Vol. 4(1).

Bowser, S. E. In preparation, Non-Cognitive Factors as Screening Tests for an Action-Research Project.

Fisher, R. H. (1936) "Use of Multiple Measurements on Taxonomic Problems". Annals of Eugenics, Vol VII, p. 179.

Johnson, P. O. (1949) Statistical Methods of Research. New York, Prentice Hall.

Mahalanobis, P. C. (1936) On the generalized distance in statistics. Proceedings of the National Institute of Science (India), Vol. 12, 49-55.

Moonan, W. J. (1972) ABCD: A Bayesian technique for making discriminations with qualitative variables. Proceedings of the 1972 Military Testing Association Conference.

Moonan, W. J. & Covhen, M. H. (1972) MIX: Mixing Probabilities from Continuous and Categorical Data for Classification Analysis. Fact Document No. 26, Navy Personnel and Training Research Laboratory.

Moonan, W. J. & Bowser, S. E. (1973) CHAROSEL Synopsis. Fact Document No. 27, Navy Personnel and Training Research Laboratory, San Diego.

Rao, C. R. (1952) Advanced Statistical Methods in Biometric Research. New York, John Wiley and Sons.

Rozeboom, J. H. (1966) The Foundations of the Theory of Prediction. Homewood, Ill.: The Dorsey Press.

Wald, A. (1944) On the statistical problem arising in the classification of an individual into one of two groups. Annals of Mathematical Statistics, 15, 145-162.

TABLE I[1]

METHODOLOGY APPLICATION

		DEPENDENT VARIABLES	
		METRICAL CRITERIA	CATEGORICAL CRITERIA
Independent Variables	METRICAL PREDICTORS	Regression Analysis [Accretion & Deletion]* Type A — Analysis of Covariance	Discriminant Analysis [Mahalanobis D^2] Type C
	CATEGORICAL PREDICTORS	Analysis of Variance [Test of Hypothesis]* Type B	Information Theory Order Statistics [CHAROSEL]* Type D

* Variable selection techniques

[1] This table derived from Rozeboom (1966, p.543) and Lubin (1950)

TABLE II

CHAROSEL APPLICATIONS

Problem Name	Criterion Categories	Number and Types of Predictors	Training Sample	Number of Predictors Used	Objective Function Value	Percent Classification Accuracy
Classification Prior to Enlistment	(2) A. Recommended for Reenlistment B. Not Recommended	(140) A. Biographical B. Self-Evaluation C. Opinion	(320) Naval Recruits	(7) (19) (95)	1.38476 1.11748 1.00000	95.6 98.4 100.0
Odds for Effectiveness Study	(2) A. Reenlistment B. Did not reenlist	(185) A. Biographical B. Self-evaluation C. Opinion	(448) Naval Recruits	(10) (43) (116)	1.24974 1.07392 1.03071	71.43 90.18 95.98
Skid Row Employment	(2) A. Completed Project B. Dropped Out	(75) A. Biographical B. Self-evaluation	(179) Skid Row Welfare Recipients	(9) (22) (42)	1.42186 1.16650 1.06096	84.4 93.3 97.8
Target Classification	(3) A. U.S. Subs B. Foreign Subs C. Surface Ships	(19) Sonar Clues	(91) Sonar Contacts	(1) (4) (19)	1.51500 1.07334 1.07334	70.3 95.6 95.6
Drug Counselors Selection	(3) A. Good + Trng B. Bad + Trng C. No Trng	(100) Flexibility Questionnaire	(81) Drug Counselors	(1) (12) (70)	2.40146 1.14078 1.01946	45.7 91.4 98.8

Table II continued on following page

TABLE II (Continued)

Problem Name	Criterion Categories	Number and Types of Predictors	Training Sample	Number of Predictors Used	Objective Function Value	Percent Classification Accuracy
Prediction of Drug Users	(3) A. Hard Drug Users B. Recreational Drug Users C. Non-Users	(139) A. Biographical B. Opinion	(362) Naval Recruits	(5) (18)	1.90207 1.72074	61.6 68.2
Recruiter Selection	(2) A. Effective Recruiters B. Ineffective Recruiters	(399) Strong Vocational Blank	(172) Recruiters	(7) (43) (68)	1.42313 1.20563 1.11167	79.7 90.7 95.4
Reserve Retention	(2) A. High Satisfaction B. Low Satisfaction	(151) A. Biographical B. Self-evaluation C. Opinion	(225) Naval Reservists	(5) (42)	1.18733 1.04840	91.6 98.2
Identification of Gifted Children	(2) A. Gifted Children B. Non-Gifted Children	(51) Behavioral Characteristics Questionnaire	(49) Kindergarten School Children	(1) (4) (30)	1.78957 1.21067 1.00000	65.3 83.7 100.0
Skyjacker Identification	(2) A. Skyjacker B. Non-Skyjacker	(10) Available Data	(475) Airline Passengers & Skyjackers	(1) (7)	1.24378 1.18300	89.9 93.1

GROWING DEMANDS ON HUMAN RESOURCES: A VIEW FROM INDUSTRY

Thomas C. Tuttle
Westinghouse Behavioral Safety Center
Columbia, Maryland

Not too many years ago, when I was fresh out of college serving my tour of duty in the Army, it was customary to speak of everything beyond the scope of the Defense Department as "the outside." In those days, those of us who did not intend to make a career out of the service would talk of how things would be different once we got out. Now that I'm standing on the outside, looking at the theme of this conference, I'm not sure that the differences are as great as I had imagined. Certainly, with regard to the topic "Growing Demands on Human Resources" there is considerable similarity. Therefore, I welcome this opportunity to look at the situation from the point of view of industry and to discuss some possible behavioral science responses to these new and increasing demands on human resources.

Perhaps a more accurate way to express the view from industry would be to use the expression "Growing Demands on Human Systems." The recent HEW Special Task Force concerning Work in America, (1972), pointed out that in many areas corporate responsibility is being expanded into areas that traditionally have received little industry attention. Increasingly the corporate balance sheet is being asked by society through legislation and social pressure to reflect some of the costs of doing business that the larger society has had to bear. This trend is reflected in the increasing concern of organizations with developing systems for human resource accounting.

There are a number of areas in which this increasing responsibility is occurring: (1) worker health, (2) worker safety, (3) quality of work, (4) environmental quality, (5) employment of women, blacks, and other minority groups. Each of these areas of corporate responsibility has generated demands on organizations to develop structures, policies, and methodologies to deal with these new problems. Thus, not only are there growing demands on the human resources in organizations, there are increased demands on the organizational systems themselves. Unless these systems become more adaptive, it will be difficult for them to make most effective use of their human resources.

In this paper I would like to describe two areas of increasing corporate responsibility in some detail, and discuss a long-range behavioral approach to these problems being developed and tested by the Westinghouse Behavioral Safety Center. The two areas to be discussed are (1)the quality of work life, and a closely related area of concern, (2) occupatinnal safety. Both of these areas are receiving increased attention by agencies in the Federal Government, such as the National Institute of Occupational Safety and Health, the Occupational Safety and Health Administration and the National Commission on Productivity.

Quality of Work

"Quality of Work" or "Quality of Work Life" are terms that encompass efforts to redesign jobs and restructure organizations to provide increased positive outcomes for workers, unions and management. The concept as defined by the National Commission On Productivity (Mills, 1973) includes all activities associated with such terms as:

job enrichment
job satisfaction
work motivation
job redesign
humanization of work
worker participation
organizational development
socio-technical development

Anyone who reads his local newspaper or the weekly news magazines is familiar with press reports concerning the declining quality of work. The terms *Changing Work Ethic, Blue-Collar Blues* and *White-Collar Woes,* have become part of our vocabulary. One frequently cited illustration of this problem is the situation at the General Motors Assembly Plant in Lordstown, Ohio. This plant, equipment and work procedures were designed as a model of efficiency. However, it neglected one crucially important consideration, the effects of this engineering marvel on the workers. The work force at the new Lordstown plant was predominantly young and not inclined to accept the regimen imposed by a super-automated assembly line. In expressing their discontent to this high speed production philosophy, the workers "responded with absenteeism, sabotage, and finally a bitter twenty-two day strike." (Jenkins, 1973). Similar reactions have been reported in varying degrees across the country.

Workers entering the work force today are characterized by higher education, higher expectations, and the demand for some sense of control over decisions that effect their lives. As a result, new workers are less willing to be bossed around and to perform monotonous, dead-end jobs. As one personnel expert stated,

> "It is becoming increasingly clear that the employees entering the world of business are unwilling to accept the system of unanswered questions, half-baked opportunities, and half-assed jobs..... They will no longer unquestioningly accept the organizational party line and will no longer stand idly by while their business lives are controlled and manipulated from above." (Jenkins, 1973, p. 78-79)

The theme of worker alienation in industry has even been picked up in a recent Army recruiting poster that says coaxingly, "If your job puts you to sleep, try one of ours."

The problem, however, goes beyond simply dissatisfaction with jobs. Workers want involvement and a share of the responsibility for decision making. Some propose job enrichment or job enlargement as a solution. Others argue that the problem goes much beyond the rearrangement of work tasks, and involves the whole structure of organizations. This latter view is one expressed in the following quote from a writer advocating "Democracy in the Factory."

> ".....the basic problem is not a faulty arrangement of jobs, it is a question of faulty power patterns; the arrangements of jobs are only surface symptoms. To be sure, any job reshuffling that gives workers increased control is a step on the road to democracy. And the assembly-line worker who sees the time span of his job increased from, say, forty to seventy seconds, and his task widened from tightening one bolt to tightening two bolts will doubtless appreciate the change. But strictly speaking, only those organizational systems that regard employee influence as a healthy and desirable phenomenon, and that are planned from the ground up to nurture and encourage such influence, can be called democratic." (Jenkins, 1973, p. 83).

A comprehensive approach to improving the quality of work must go beyond a concern with job and task considerations, although they are very important, and must consider more fundamental organizational characteristics as well.

Occupational Safety

The problems of occupational safety have been highlighted by the passage in 1970 of the Williams-Steiger Occupational Safety and Health Act, and the 1972 President's Report on Occupational Safety and Health. The President's Report on Occupational Safety and Health (1972) documents the fact that between 1961 and 1970 the injury

frequency rate for manufacturing industries (i.e. the number of disabling injuries per million employee-hours worked) has increased from 11.8 to 15.2, an increase of approximately 29% in ten years. This report also presents data resulting from safety inspections conducted in 1971 under the provisions of the Williams-Steiger Occupational Safety and Health Act. Out of 13,174 inspections conducted there were 31,053 reported violations. Although many of these reported violations were not in the serious category, they provide some indication of the general climate of the organizations inspected with regard to safety philosophy.

In terms of the number of employed persons who experience health or safety problems related to employment there are no firm statistics. Nevertheless, the National Safety Council (President's Report on Occupational Safety and Health, 1972) estimates that of the 80 million employed persons in the civilian labor force, over 14,000 suffer fatal accidents and 2.2 million suffer disabling injuries each year. It should be emphasized that these numbers are only estimates. They do not attempt to account for the unknown number of non-disabling accidents and illnesses resulting from exposure to hazards.

An estimate of the tangible costs resulting from industrial accidents can be derived from the expenses incurred by employers in the form of workman's compensation. The Social Security Administration estimates that the total cost of the Workman's Compensation to employees in 1970 was 4.82 billion dollars. (President's Report on Occupational Safety and Health, 1972) Even this figure is an underestimate of the actual cost, due to the fact that many workers are not covered by the act. In addition the figure does not include costs due to retraining, rehabilitation necessitated by disabling accidents, as well as less tangible costs such as emotional disturbance, and disturbances in job performance or home life. These indices

substantiate the fact that accidents and illnesses attributable to the work environment remain a serious problem.

Relationship between Quality of Work and Safety

Occupational Safety and quality of work life are both intimately related in terms of their impact on the employee and in terms of organizational factors that effect them. Certainly a worker who must face an extremely hazardous job on a daily basis cannot be considered to have high quality work life. Furthermore, organizations that are attentive to the safety hazards effecting workers are also likely to pay attention to the effects of other organizational conditions on employee satisfaction and performance. Therefore, occupational safety concerns can be considered to be one aspect of the overall quality of work. Organizational factors that impact positively on quality of work criteria such as productivity, satisfaction, absenteeism, and turnover can also be expected to have a positive impact on safety criteria.

Organizational Development Approach

The most popular, recent approach for dealing with such human problems in industrial organizations has been that of the organizational development specialists. Many of these so-called OD specialists have claimed considerable success in dealing with problems of worker alienation, declining productivity, high absenteeism and rapid turnover. The size of the consulting fees commanded by some of these OD "super-stars" represents one measure of their success. Despite the fact that OD specialists have undoubtedly had a positive impact on a number of organizations, the approach as a pragmatic answer to problems throughout industry seems to suffer from two major weaknesses.

First, in many ways the practice of OD is more of an "art" than a science at this point. The OD specialist typically collects two types of information, subjective impressions from observations and interviews, and objective survey data. From this data, diagnoses are made and recommendations for action formulated, often in consultation with organizational personnel. As a result of the subjectivity of much of the data and the diagnostic process, itself, it is difficult for a specialist to communicate to someone else what he has done. As a result the approach lacks rigor and replicability.

A second weakness is that there just aren't enough OD specialists available to deal with the problems facing the industrial community. The process is quite time-consuming from the point of view of the consultant, therefore, one individual can consult with a relatively small number of organizations in a period of a year. Furthermore, training of effective OD consultants is complicated by the vagueness surrounding the behaviors required to successfully operate in the organizational setting. As a result it seems unlikely that Organizational Development specialists can be trained in adequate numbers to meet the demands placed on them.

Westinghouse Behavioral Safety Center Approach

How then will organizations be able to cope with the growing demands placed upon them? What methodologies can be used to improve the quality of work life for management, unions, and employees? Obviously there are many answers to such a question. Just as obviously, there are probably none that are entirely satisfactory. Nevertheless, I would like to discuss an approach currently being developed by the Westinghouse Behavioral Safety Center which we feel has considerable promise.

This approach grew out of in-house research as well as research conducted under contract to the National Institute of Occupational Safety and Health (NIOSH).[1] The approach recognizes that organizations, whether they are military, private industry, government or educational, are complex. Thus, any approach must consider a wide range of variables in order to include those variables that effect quality of work or safety outcomes. The first step of this research program involved development of a set of behavioral guidelines by surveying research literature in four areas of psychology: (1) organizational psychology, (2) behavior modification, (3) training and (4) engineering psychology. These guidelines reflect the state of the art in each of these four areas with primary regard for variables that have implications for accident reduction. However, as discussed further below, the guidelines have broader applicability for other organizational outcomes. Each guideline consists of a short definition, followed by a 1-3 page description of the supporting research literature and the implications of the guidelines for organizational functioning. Approximately 80 quidelines were defined, distributed approximately equally across the four research areas. The quidelines were written primarily for the safety professionals, rather than for behavioral scientists, thus they may appear to be somewhat oversimplified.

---

1 Westinghouse Behavioral Safety Center contracts with NIOSH include:
HSM 99-72-27 Psychological Behavioral Strategy for Accident Control

HSM 99-73-61 Implementation of Psychological Behavioral Strategy for Accident Reduction

Examples of selected guideline definitions are presented below. Included are two guideline definitions from each of the four research areas. The complete outlines giving the titles for all the guidelines in each of the four research areas appear in Appendix A.

1. Organizational Psychology (Dachler and Schneider, 1973)
   a. Maintaining group norms. Group pressure is an effective means for maintaining behavior in accordance with group norms that encourage safety. Organizations can enhance the effectiveness of group pressure by (1) creating work groups that are attractive to workers, (2) promoting cooperation among work groups and (3) by encouraging social contact within the workplace.
   b. Job design. Jobs should be designed to enhance worker interest and involvement. Jobs which are designed to emphasize intrinsic factors (e.g., recognition, variety, autonomy, responsibility) tend to contribute to higher worker satisfaction and improved work quality

2. Behavior Modification (McIntire and White, 1973)

   a. Social reinforcement - supervisory. Social reactions from supervisors such as praise, attention, ridicule, laughter, or facial gestures are effective consequences of behavior. They can be used to increase the frequency of behavior which promotes safety and to decrease the frequency of behavior which contributes to accidents.

   b. Non-contingent reinforcement. If reinforcers are not given in response to specific behaviors and according to an explicit and predictable pattern, they cannot be expected to increase the likelihood of desired safety behavior.

3. Training (Goldstein, 1973)

   a. Feedback. Knowledge of results serves a motivational function and a guidance function. As a general rule, training aids should be designed to provide maximum and immediate feedback to the trainee as to the adequacy of his safety performance. It is further important to determine that the feedback is being used by the trainee. For learning purposes it is useless to present the trainee with feedback information which he cannot understand.

   b. Overlearning. Tasks or procedures that may have to be performed in emergencies or other stressful situations should be practiced until they become overlearned. Such overlearning increases the resistance of task performance to disruption or interference under stress.

4. Engineering (Grether, 1973)

   a. Job safety design. Organizations should ensure that jobs are designed to minimize human-initiated errors and accidents. Specifically, jobs should be comprised of related sets of task activities that are compatible with each other in terms of task sequencing, time utilization, location, and function. The tasks should be logically grouped and allocated to either individuals or work teams.

   b. Work/rest cycle. In establishing policies relating to work/rest period scheduling, organizations should ensure that (1) the work duration does not exceed four consecutive hours, (2) the duration of rest periods permits physiological and cognitive recovery from task demands, (3) rest periods are scheduled throughout the day or work period in accordance with the level of environmental or task demands, and (4) scheduling and duration of rest periods allowed for social contact with co-workers.

The full-set of behavioral guidelines represents the knowledge base for the long-term approach. In addition the guidelines provide some immediate payoff since they give the practitioner a summary in relatively simple terms, of some relevant behavioral science principles. This summary, although not an answer to all his problems, can be used to assist him in thinking about safety problems, instructional problems, or any of a wide range of human resources problems he may face.

Another key ingredient of the approach is represented by the Organizational Performance Intervention Model (Appendix B). This model serves to identify the major categories of variables assumed to be relevant to quality of work and safety criteria, and provides a means for integrating a wide range of variables into a single approach. It also serves to make explicit some of the assumptions underlying this research program.

During the development of this model, it became obvious that the original assumption of the approach was valid. This assumption held that the variables which effect the level of safety in the organization, also effect other outcomes such as job satisfaction, productivity and overall "quality of work." Thus, the guidelines developed for safety,can with very little modification, be considered to be guidelines for improving overall quality of work. For example, job design is an important variable with regard to safety as well as for quality of work criteria such as job satisfaction or productivity. Jobs that are composed of imcompatible tasks or which lead to conditions of information overload are likely also to lead to high levels of stress, low satisfaction and increased likelihood of accidents. Despite the fact that the guidelines were developed in the context of occupational safety, their generality is not restricted to safety concerns.

The model and the set of behavioral guidelines form a background for the Westinghouse long-range,quality of work research program. This background provides a firm base from which to launch organizational intervention programs. The intervention strategy is characterized by three phases: (1) Organizational Diagnoses, (2) Implementation and (3) Evaluation.

The organizational diagnostic procedure attempts to avoid the lack of rigor inherent in the typical organizational development type of diagnosis. One aspect of the diagnostic procedure includes a diagnostic inventory to be completed by employees. The diagnostic inventory is based on the organizational model discussed earlier and includes items designed to measure the following sets of variables: (1) organizational characteristics (e.g., structure, decision making, policies, etc.) (2) supervisor and work group characteristics (e.g., supervisory style and competence, competition, co-operation, etc.) (3) job and task characteristics (e.g., role ambiguity, information overload, variety, etc.) (4) individual attitudes (e.g., dimensions of satisfaction) (5) characteristics of the work environment (e.g., lighting, temperature, adequacy of supplies, etc.). In addition to the diagnostic inventory the diagnostic procedure also includes provision for a company records audit. This procedure will yield criterion information on variables such as: (1) productivity, (2) absenteeism, (3) turnover, (4) accidents, and (5) grievances. The criterion information is needed in order to establish statistical relationships between organizational process variables and individual and organizational outcomes. Furthermore it will provide baseline data for the evaluation phase.

Implementation

The overall diagnosis yields an organizational profile along a number of relevant dimensions. This is useful to the extent that it allows for recommendations for action that are tailored to the needs of the organization. The mechanism for bridging the gap between

the diagnosis and intervention is based on a set of action modules and a tracing procedure for linking the action modules to the diagnostic procedure.

The action modules are sets of actions that an organization can take to accomplish some specific objective. These modules are based on the previously discussed behavioral guidelines, and represent essentially a translation of the guidelines into action terms. Each module is oriented toward some specific objective such as implementation of a supervisory reward system, development of a safety training program, implementation of a behavioral record keeping system, or job redesign. For example, the job redesign action module would combine several guidelines from the Organizational Psychology and Engineering Psychology sections. Based on these guidelines, specific actions that should be considered in redesigning a job would be stated. Examples of actions might be: (1) analyze worker abilities and interests; (2) determine function or functions to be performed, (3) identify tasks needed to perform required functions, (4) obtain worker suggestions concerning grouping of tasks into positions, (5) form groups of compatible tasks having sufficient variety to maintain worker interest, (6) evaluate worker reactions to redesigned job.

The next step in the research program involves development of a tracing system for linking the action modules to the diagnostic procedure. There are a number of possible approaches to this problem. One procedure that seems promising involves ratings by experts of the relevance of each action module for each of the items in the diagnostic inventory. This procedure would yield a vector of weights for each diagnostic item corresponding to the action

module. Once an organizational profile is obtained on the diagnostic inventory the item ratings would be combined with the action module weights and summed across diagnostic items. This would produce a score for each action module, making it possible to rank the modules in terms of their relevance for the organization analyzed. Depending on a consideration of cost-factors the relevance of various action strategies, and the interrelationships among the action modules, decisions can be made with regard to the modules to be combined to form a program.

Evaluation

Once the action program is implemented, the results will be evaluated in terms of cost and behavioral criteria related to stated organizational goals. Initially, this will be done in an effort to validate the methodology. However, evaluation is essential also due to the fact that organizations making or contemplating changes must conduct careful evaluations in order to derive cost-effectiveness data. Ultimately the method will be successful only if it leads to cost-effective organizational change.

Concluding Statement

The methodology described represents an ambitious undertaking which is now in its preliminary stages. Although the approach remains to be validated in organizational settings, it offers a number of advantages over conventional organizational development techniques. The method is explicit, it is systematic, and it offers some degree of rigor. Ultimately it should be possible to provide the capability for organizations to diagnose themselves and to make changes to meet human resource challenges without the intervention of an outside party. This is the ultimate goal of the procedure.

REFERENCES

Goldstein, I. Training guidelines, In Tuttle, T. et al.(Eds.) Psychological behavioral strategy for accident control: Development of behavioral safety guidelines, Final report for NIOSH Contract No. HSM 99-72-27, Columbia, Maryland: Westinghouse Behavioral Safety Center, 1973.

Grether, C. Engineering psychology guidelines. In Tuttle, T. et al. (Eds.) Psychological behavioral strategy for accident control: Development of behavioral safety guidelines, Final report for NIOSH Contract No. HSM 99-72-27, Columbia, Maryland: Westinghouse Behavioral Safety Center, 1973.

Jenkins, D. Democracy in the factory, The Atlantic Monthly, 1973 (April), 78-83.

McIntire, R. and White, J. Behavior modification guidelines. In Tuttle, T. et al. (Eds.) Psychological behavioral strategy for accident control: Development of behavioral safety guidelines, Final report for NIOSH Contract No. HSM 99-72-27, Columbia, Maryland: Westinghouse Behavioral Safety Center, 1973.

Mills, T. The quality of work program of the national commission on productivity: A statement on intent. Unpublished paper, 1973.

President's report on occupational safety and health. New York: Common Clearing House, 1972.

Dachler, P. and Schneider, B. Organizational psychology guidelines. Working paper submitted to Westinghouse Behavioral Safety Center, 1973.

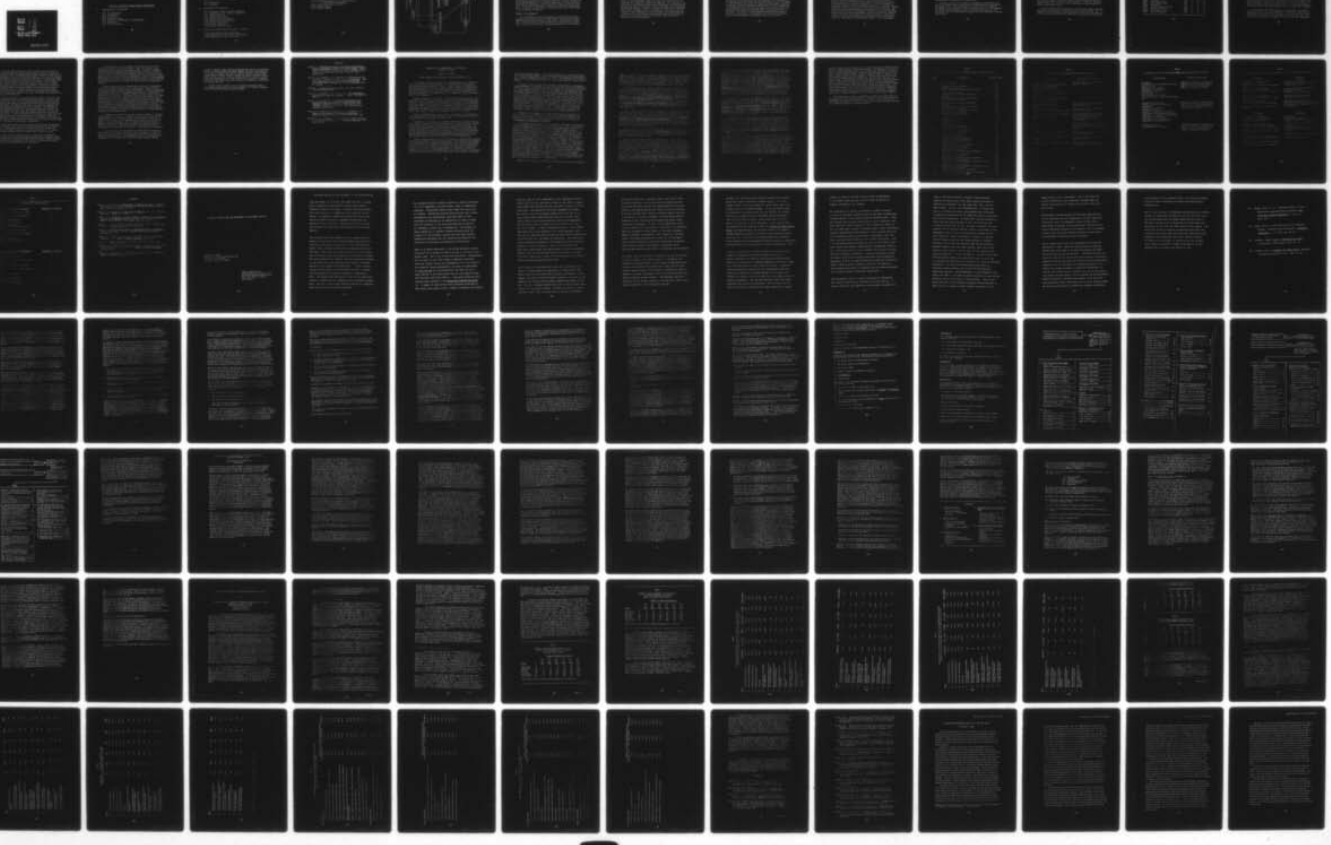
Tuttle, T., Grether, C. and Liggett, W., (Eds.) Psychological behavioral strategy for accident control: Development of behavioral safety guidelines, Final report for NIOSH Contract No. HSM 99-72-27, Columbia, Maryland: Westinghouse Behavioral Safety Center, 1973.

Work in America, Report of a special task force to the Secretary of Health, Education and Welfare, Cambridge, Massachusetts: The MIT Press, 1972.

AD-A040 773 AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TEX F/G 5/9
PROCEEDINGS, ANNUAL CONFERENCE OF THE MILITARY TESTING ASSOCIAT--ETC(U)
NOV 73

UNCLASSIFIED NL

7 OF 9
ADA
040773



A. Organizational Objectives and Policies

O-1 Extra organizational environment
O-2 Organizational objectives
O-3 Organizational decision making
O-4 Organizational communication
O-5 Attitudes toward safety
O-6 Reward policy

B. Job/Individual Compatibility

O-7 Job design
O-8 Worker expectations
O-9 Job assignment
O-10 Job stress

C. Group and Interpersonal Processes

O-11 Establishing group norms
O-12 Maintaining group norms
O-13 Intergroup competition
O-14 Intergroup co-operation

D. Leadership/Supervision

O-15 Sources of influence
O-16 Leadership style
O-17 Impact on safety motivation

E. Organizational/Individual Feedback

O-18 Criterion development
O-19 Performance counseling

Outline of Behavior Modification Guidelines

A. Consequences of Behavior

B-1 Reinforcement
B-2 Punishment
B-3 Avoidance
B-4 Response cost
B-5 Social reinforcement - supervisory
B-6 Modeling
B-7 Peer pressure

B. Scheduling of Consequences

B-8 Immediacy
B-9 Consistency
B-10 Shaping
B-11 Partial reinforcement
B-12 Extinction
B-13 Compensatory reinforcement
B-14 Stimulus control

C. Pitfalls in scheduling consequences

B-15 Non-contingent reinforcement
B-16 Non-contingent punishment
B-17 Reinforcement opportunity

D. Implementation - General Considerations

B-18 Practice
B-19 Record keeping

Outline for Training Guidelines

A. Transfer of Training

T-1 Training and testing similarity
T-2 Transfer and understanding
T-3 Transfer and practice on the original task

B. Motivation and Trainee Behavior

T-4 Motivation
T-5 Feedback

C. Conditions of Practice

T-6 Guidance toward correct responses
T-7 Opportunity for correct responses
T-8 Mental practice
T-9 Learning to learn
T-10 Massed-spaced practice
T-11 Part-whole task practice
T-12 Stages of training
T-13 Verbal labelling
T-14 Use of learning time

D. Special Considerations Related to Training

T-15 Overlearning and emergencies
T-16 Learning and individual differences
T-17 Membership in the work group

Outlines For Engineering Psychology Guidelines

A. Work Organization

E-1 Function allocation
E-2 Shift work
E-3 Job safety design
E-4 Work/rest cycle

B. Job Characteristics and Demands

E-5 Arousal and stress
E-6 Fatigue
E-7 Work pace

C. Workplace Design and Layout

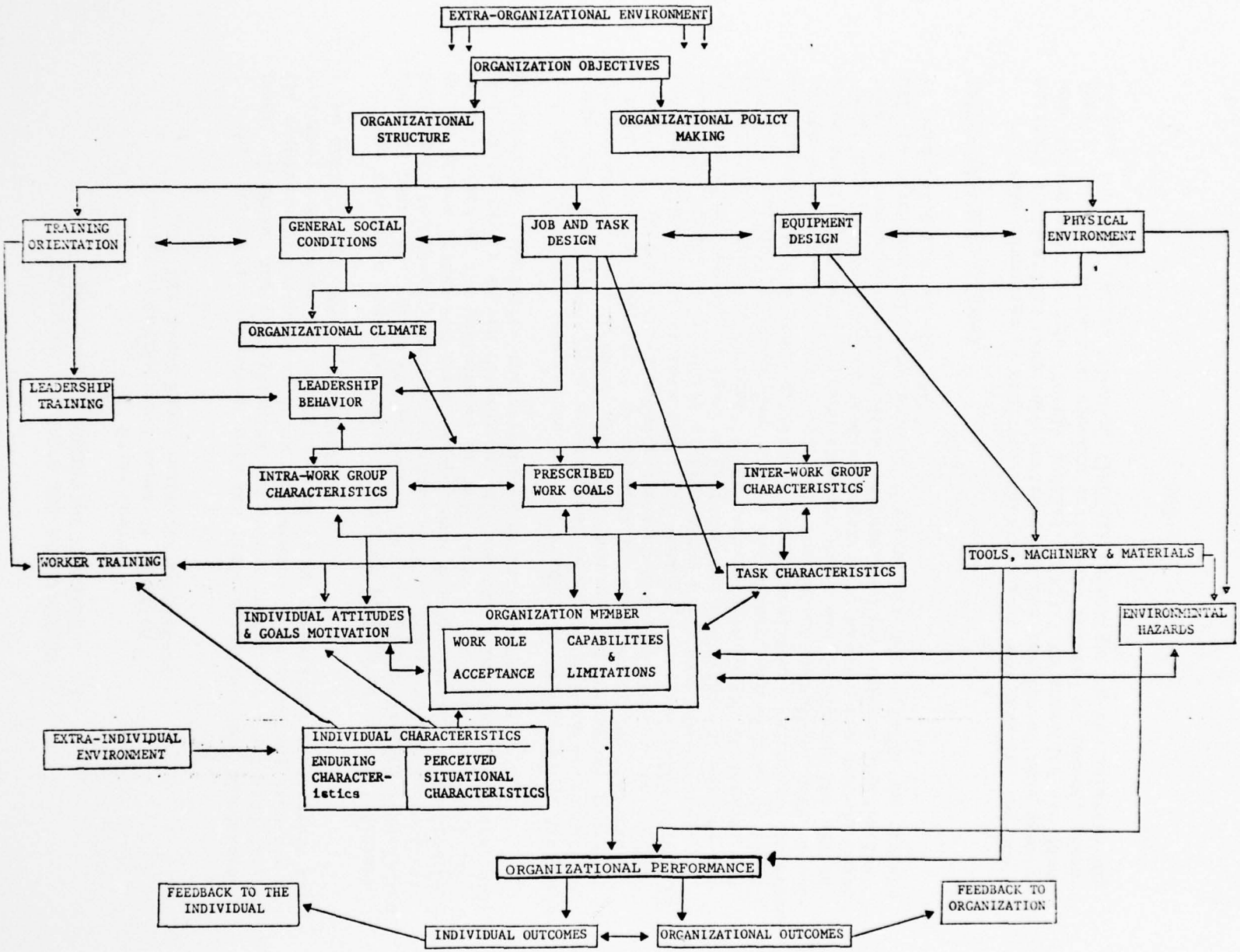
E-8 Workplace illumination
E-9 Anthropometric design
E-10 Warning signals
E-11 Equipment and control-display arrangement
E-12 Location of visual displays
E-13 Location and protection of control devices

D. Task Characteristics and Demands

E-14 Stimulus response (S-R) uncertainty
E-15 Task information processing load
E-16 Skilled motor performance
E-17 Redundant information
E-18 Task performance/memory aids

E. Control-Display Characteristics

E-19 Stimulus-response (S-R) compatibility
(population stereotypes)
E-20 Coding and identification
E-21 Display design
E-22 Control design



Appendix B WESTINGHOUSE MODEL OF ORGANIZATIONAL PERFORMANCE - INTERVENTION
(Adapted from Tuttle et al. 1973)

REVIEW OF AIR FORCE JOB SATISFACTION AND CAREER DEVELOPMENT RESEARCH

R. Bruce Gould
Occupational Research Division
Air Force Human Resources Laboratory

The zero draft environment with its reduced aptitude and educational levels of airman inputs, strengths and budget reductions, increased training requirements through technological advances, and resulting emphasis on personnel utilization and retention provide an operational requirement for a job satisfaction research program.

The Air Force Human Resources Laboratory program of job satisfaction research began in 1971 as was announced to the Military Testing Association Conference that year (Gould & Christal, 1971). The effort is an outgrowth of the United States Air Force Occupational Survey Program, which provides data obtained at the performance or task level. The task-level data are central to this research program and differentiate it from the theoretical research which dominates the psychological literature.

This paper will present the ongoing studies conducted by the Air Force Human Resources Laboratory's Occupational Research Division in support of the job satisfaction research plan which has evolved during the two-year period. The long-range goal of the plan is retention of job qualified personnel. To this end, the basic steps of the plan are to: (a) define the dimensions of job satisfaction, (b) measure the satisfaction levels of these dimensions, (c) identify problem areas which have the greatest potential for improvement through satisfaction research, (d) measure the effects which specific changes in job content have on job attitudes, and (e) implement job reengineering actions and measure the effects on job attitudes and on the final criterion, career decisions. Although the steps in the long range program are logically sequential in nature, current projects studying segments of each step are under way to provide the needed technology for longitudinal studies required to support the overall objective. An overview of the studies' purposes and preliminary implications will be discussed, although some of the projects are still in the process of data collection or analysis.

Review of the Literature

To take advantage of previous research, some 400 civilian and 284 military publications have been reviewed and the results of this massive undertaking are in press (Tuttle & Hazel, 1973). To summarize the findings: (a) Job satisfaction appears definitely related to tenure, absenteeism, and sick calls, however, findings investigating the relationship

between satisfaction and performance are inconclusive; (b) Job satisfaction is multidimensional, with satisfaction and dissatisfaction representing polar ends of a continuous attribute rather than two separate attributes as presented by Herzberg (1966) and his associates (Whitsett & Winslow, 1967); (c) Job satisfaction is a function of the interaction between work factors and personnel needs, that is, the extent to which expectations or needs of the individual are fulfilled by his work environment; (d) Most studies have dealt with subjective reports of attitudes toward jobs -- very few have dealt with specific job content; (e) There have been few dynamic studies relating changes in job content to changes in attitudes; (f) The greatest volume of literature deals with overlapping, frequently conflicting, theories of job satisfaction which are probably all oversimplifications of the relationships with which applied researchers must concern themselves.

Past studies tend to support the view that more practical applications of satisfaction research will result from studying individuals' satisfaction with specific facets of jobs rather than using a global measure of job satisfaction. Although there are several instruments designed to measure satisfaction with specific aspects of jobs, notably the Minnesota Satisfaction Questionnaire (Weiss, Dawis, England, & Loftquist, 1967) and the Job Description Index (Smith, Kendall, & Hulin, 1969), none were found which covered all major aspects of military work environments. This necessitated development of a suitable instrument, applicable to all Air Force specialties.

Development of the USAF Occupational Attitude Inventory

Based on the literature review, a list of presumed attitude content areas or facets of satisfaction was compiled and each facet briefly defined and reviewed by behavioral scientists familiar with the military work environment. Thirty-four work dimensions (shown in Table 1) resulted, as compared to 20 in the Minnesota Satisfaction Questionnaire and 5 in the Job Description Index. Items were then written by three behavioral scientists for each defined facet resulting in an original pool of 440 items. Using 6 independent judges, the items were sorted into the 34 dimensions and reviewed for clarity and redundancy. For selection of the final items, those showing classification disagreement were inspected for ambiguity and revised if possible. Obviously redundant items and remaining ambiguous items were removed, resulting in a final pool of 348 items. Table 2 shows an example of five of the specific items included in the inventory. Full documentation of the development is nearing completion. It suggested that the extensive item pool and literature review will be useful to other military services in developing their own service-specific job attitude instruments.

TABLE 1. WORK DIMENSIONS OF THE OCCUPATIONAL ATTITUDE INVENTORY

Achievement	Personal Growth and Development	Service to Others
Activity	Job Design	Social Status
Air Force and unit policies and practices	Optional Social Contact	Sufficiency of Training
Assignment locality	Pay and Benefits	Supervision Received - Human Relations
Authority	Physical Work	Supervision Received - Technical
Co-Workers	Recognition	Performance Evaluation
Creativity	Required Social Contact	Job Change
Importance	Responsibility	Tools, Equipment, and Supplies
Independence	Physical Safety	Utilization
Knowledge of Results	Economic Security	Variety
		Work Schedule

The 348 items make operational use of the inventory prohibitive, because of examinee attention span and administration time constraints, but they provide a useful experimental instrument. The experimental Occupational Attitude Inventory has been administered to a random sample of 5,800 airmen and the returns are currently being analyzed. An 8-point rating scale with point descriptors ranging from Extremely Dissatisfied to Extremely Satisfied was selected for inventory use. The primary goal of this analysis is to reduce the instrument to the smallest subset of items which will measure the attitude domain variance captured by all 348 items. To do this, a computer program titled VARSEL, Variable Selection in the Absence of an External Criterion, has been developed. Using an accretion method, the program starts with one item and

iteratively selects items having the greatest unique variance of all the remaining nonselected items until the combined pool of selected items can predict the variance measured by each nonselected item. The computational method uses a multiple linear regression technique and has options for establishing cutoff points. These options consider item reliabilities and permit investigator interaction with the item selection process. Preliminary findings from the analyses suggest that approximately 150 to 170 items will remain in the final inventory, and that these items measure about 40 internally consistent job attitude dimensions. These observations are supported by profile, regression, and factor analyses.

TABLE 2. EXAMPLES OF SPECIFIC JOB SATISFACTION ITEMS RELATING TO SUPERVISION

1 = Extremely Dissatisfied 2 = Very Dissatisfied 3 = Moderately Dissatisfied 4 = Slightly Dissatisfied	5 = Slightly Satisfied 6 = Moderately Satisfied 7 = Very Satisfied 8 = Extremely Satisfied	Write your answer in this column
1. The extent to which supervisor keeps you informed		1. _____
2. Supervisor fairness in rating subordinates		2. _____
3. Way supervisor handles technical problems		3. _____
4. Supervisor's knowledge of how your job is done		4. _____
5. Willingness to listen to employees		5. _____

Development of an Assignment Preference Survey

An additional measurement instrument is being developed to evaluate attitudes toward assignment locations. Job satisfaction studies in specific specialties indicated that assignment location frequently plays a substantial role in attitudes towards jobs. The current assignment system only considers positive assignment preferences. Apparently, these preference ratings are often made by individuals who are relatively ignorant of the physical and environmental characteristics of the locations. The question is, what are the environmental characteristics that an individual would consider in selecting assignment preferences if they were available to him? Once these are determined, it should be possible to develop an assignment system which provides objective information on these characteristics for consideration in assigning preferences. Such an assignment

system could be expanded to consider not only a man's likes but his dislikes. It may be more important to the individual's attitude to not assign him to an area of extreme dislike than for him to get his most preferred assignment.

An Assignment Location Preference Survey has been administered to 8,000 airmen to provide two ratings:

1. Their assignment preferences for 150 unnamed CONUS locations given objective data on factors shown in Table 3.

2. Ratings of 150 corresponding base names.

Because the locations correspond to the bases rated, the extent of base environmental ignorance among airmen can be evaluated as well as the environmental factors most relevant to assignment preferences.

TABLE 3. AIRMEN - BASE ASSIGNMENT PREFERENCE SURVEY
FOR 150 MAJOR CONUS INSTALLATIONS

ENVIRONMENTAL FACTORS
Base military population
Size of community nearest base
Miles to 200,000 size city
Cost of living compared to national average
Availability educational facility (college)
Distance: mountains, desert, ocean, lakes, rivers
Average July - January temperature
Average summer - winter weather (humidity - rainfall - sunshine)
Annual snowfall

Identification of Specialties for Indepth Study

Responses of some 175,000 airmen job incumbents to occupational surveys in 150 career ladders are available for identification of specialties for indepth study of job satisfaction. The data consist of background information on job incumbents, task-level definitions of jobs, and responses to a career intent statement and two job attitude questions; i.e., how interesting the job is and how well it makes use of the respondent's talents and training.

From the inventory data, we can state that most airmen find their jobs interesting and report their talents and training are well utilized. However, there are extensive differences within and between career ladders and there is some indication that over the last few years, airman reports of dissatisfaction have been increasing. For example, in the 150 career ladders surveyed to date, 13 have been surveyed twice with 3-year intervals between surveys. During the 1967 to 1969 period, 24% of the airmen at the 3-(semi-skilled) and 23% at the 5-(journeyman) level reported their talents and training as utilized "Very Little" or "Not at All." During the resurvey period of 1970 to 1972, 42% and 37% of the 3- and 5-level airmen in these ladders reported poor utilization. The substantial increases in reported dissatisfaction may be due to any of several reasons as in deterioration of jobs or restrictions in job progression. It may just be that airmen are more willing to express negative attitudes or this may be just symptomatic of the general attitude of the younger civilian population from which our airmen are obtained. Popular civilian mediums are reporting substantial increases in dissatisfaction with work performed. Job enrichment programs are becoming the vogue in civilian industry to counter the trend.

The job attitude differences among career ladders can be illustrated by comparing the percentage of airmen reporting that their talents and training are poorly utilized. Rank ordering the ladders from highest to lowest percentage of 3-level airmen reporting poor utilization, and comparing the two extremes of that distribution (see Table 4), the range is from 63% of the 3-level Pavements Maintenance personnel to none of the Dental Laboratory Technicians (Gould, 1972). Within ladders, the percentage generally decreases as the skill level increases, especially from the 5- to 7-level. Airmen at the 7-level are fully qualified E-6 or E-7 career airmen.

Comparing the specialties at the extreme ends of the distributions, there are no readily apparent explanations for the differing utilization reports. At both ends there are ladders with high and low aptitude entrance requirements and diverse content areas, i.e., electronics, equipment repair, security, and medical ladders. This suggests that there are few universal causes of reports of dissatisfaction. Individual studies of ladders indicate that factors related to dissatisfaction are essentially unique to each career ladder.

Several examples will be cited which illustrate that specific problems are unique to each ladder and ladders must be separately studied. For example, in the Intelligence Operations ladder, it was found that over half of the tasks airmen were trained to perform in the mandatory 30-week entry-

level technical schools were never performed in the field. The technical school had combined the curriculum for Photo Interpreters and Intelligence Operators because of common elements. This had the dual effect of increasing training of nonrelevant tasks and increasing the required aptitude entry level. Aptitude requirements are set by predictability of technical school success and the photo interpreter tasks were highly demanding thus raising the aptitude required to master the tasks. The curriculum has been adjusted and the aptitude requirement lowered to the requirements of the actual tasks performed.

TABLE 4. PERCENT CLAIMING TALENTS AND TRAINING ARE UTILIZED "VERY LITTLE" OR "NOT AT ALL"

AFSC	Career Ladder	N	Skill Level 3	5	7
551X0	Pavements Maintenance	1059	63	54	15
276X0	A/C Control and Warning Tech	1689	63	51	27
461X0	Munitions Maintenance	1590	58	50	21
402X0	Aerospace Photo Sys	261	-	56	35
671X3	Accounting Disbursement	1354	58	45	N/A
611X0	Supply Services	949	53	36	12
303X3	Auto Tracking Radar Rpmn	708	51	39	20
234X0	Precision Photoprocessing	836	47	42	15
811X0	Security	3617	46	36	9
443X0G	Minuteman Missile Rpmn	1032	44	46	18
204X0	Intelligence Operations	926	43	39	20
915X0	Medical Materiel	1022	43	26	7
	.	.	.	.	.
	.	.	.	.	.
	.	.	.	.	.
811X0A	Dog Handler	559	10	24	2
903X0	Radiology	548	9	12	11
432X0	Jet Engine Mechanic	1622	9	8	6
305X3	Elect Computer Rpmn	1055	7	14	13
607X0	Loadmaster	1773	6	12	5
324X0	Precision Measuring Equip	1212	6	7	8
322X1	Weapons Control Sys	1285	5	6	3
272X0	Air Traffic Control	1518	2	5	7
982X0	Dental Laboratory	486	0	6	3

In the Automatic Tracking Radar Repairmen ladder it was found that the 37-week technical training concentrated on teaching highly sophisticated electronic tasks which were not likely to be performed by an airman until his second enlistment. There are two approaches to this problem: (a) re-engineer the lst-term airman's jobs to coincide more with those of the career airman, and hence the initial training program; or (b) provide initial training comparable to the tasks encountered in the first years, and make the advanced training available after reenlistment.

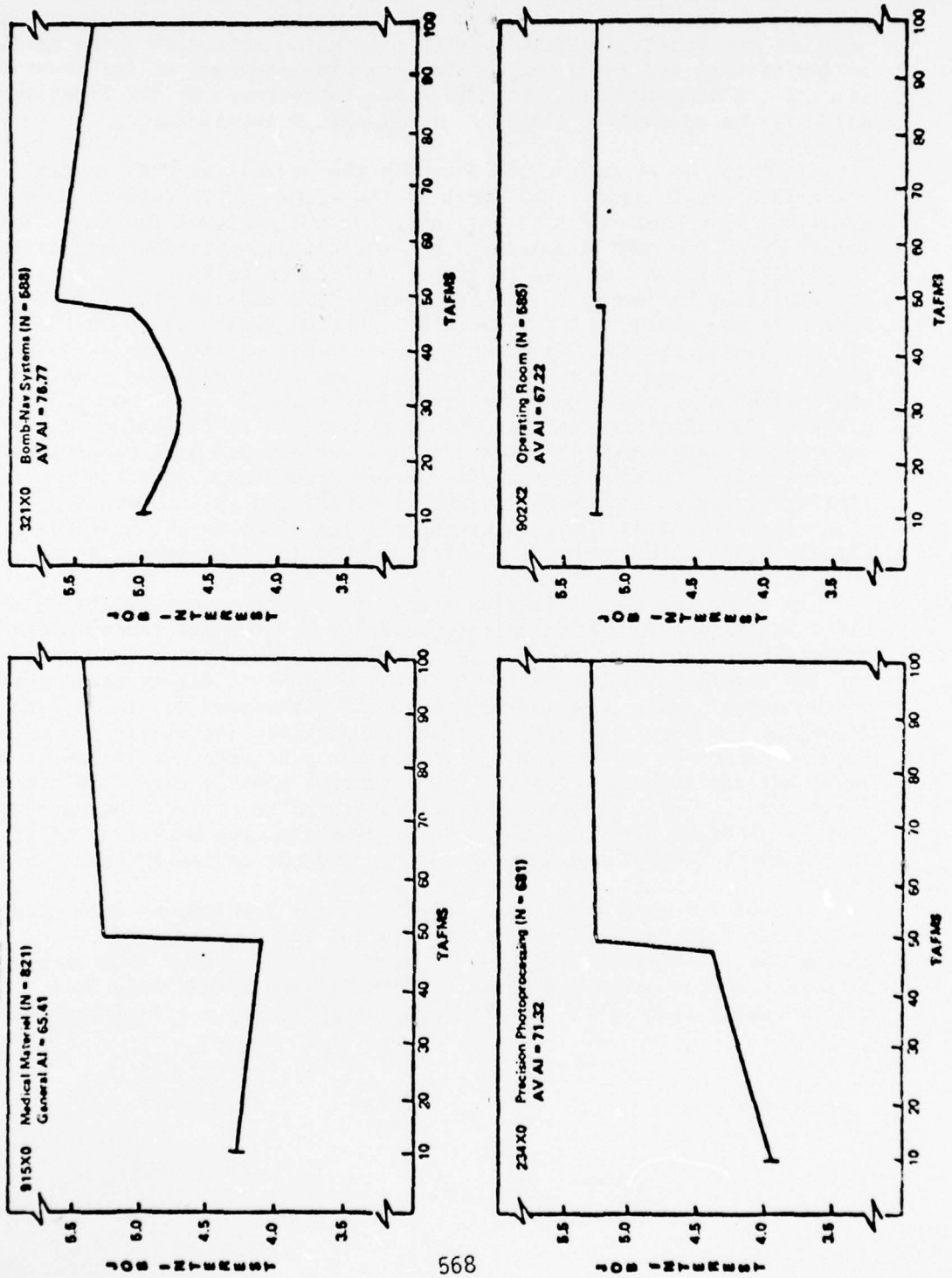
A different situation was found in the Accounting Disbursement ladder where in a single year (1968) most of the airmen input were college graduates who scored at the 95th percentile on all aptitude indices. This career field has many demanding jobs, but not in sufficient proportions to provide challenging jobs to such an influx of talent. The problem is not likely to be reencountered now that we are experiencing reduced talent inputs in the zero draft environment. Another issue raised in the study of this ladder was the impact which supervisory ability has on job attitudes. As in several Air Force specialties, this ladder is combined with other ladders in the accounting specialty at the 7-level. A study now in progress is using occupational survey techniques to determine critical supervisory tasks necessary for adequate performance as a supervisor. The results will have implications for career progression patterns and career field structures. Interim findings in other ladders indicate that adequate supervisors are important to incumbent's job attitude as was found particularly true in the Aircraft Control and Warning Technician ladder.

To date, the most extensive study of satisfaction elements in a specific ladder was conducted in the Pavements Maintenance ladder where dissatisfaction was found to be largely a function of the mundane nature of the job domain. This study is important because it demonstrated that occupational survey data followed by field interviews can be used to determine the nature of job satisfaction problems and provide remedial recommendations. However, field interviews are expensive in time and manpower and the next question is: "Can surveys also be used in place of the field interviews"? Surveys specially designed to collect the data previously obtained by interviews have been used in the Munitions Maintenance and Aircraft Control and Warning ladders and are currently being analyzed.

One of the more ambitious projects is the development of a single large data file for the study of airman job satisfaction. Thus far, the occupational survey data of 130,000 airmen in 140 career ladders has been matched with the personnel files and placed in a single data file. The file contains task-level performance, job attitude, personal background,

aptitude, and loss -- gain data. It constitutes the largest existing data file, to this investigator's knowledge, constructed for job attitude research.

TABLE 5. JOB INTEREST VS TAFMS (HOLDING APTITUDE CONSTANT)



One of the first uses made of the file will be identification of career ladders exhibiting the greatest potential return for indepth job satisfaction study. For example, Table 5 shows regression curves for four career ladders. Here a job attitude dimension, in this case job interest, is represented by the vertical (Y) axis, and months of Total Active Federal Military Service (TAFMS) is indicated on the X axis. In predicting interest, the actual predictors used are career status (1st-term or career airman), time-in-service for 1st-term airmen, service time for career airmen, aptitude, and the TAFMS values squared. Aptitude is included in the regression model for several reasons. We have data which indicate that aptitude levels are strongly related to job attitudes and these levels vary with time. Since we are here interpreting cross-sectional data in a longitudinal fashion, the effects of aptitude differences across time must be held constant.

Compare the profile of the Medical Materiel ladder in the upper left corner of Table 5 with the Operating Room Technicians in the lower right. Job interest decreases for the 1st-term Medical Materiel airmen as they progress in their career and approach the reenlistment decision point at the 48th month. The jump from the 48th to the 49th month suggests that those reporting uninteresting jobs did not reenlist. The magnitude of the jump indicates the degree of relationship between job interest and the reenlistment decision. Note the positive acceleration of the curve from the 49th to the 100th month which suggests some type of perceived job progression. A ladder exhibiting such a profile has a high likelihood of impact on positive job attitudes and career decisions through job reengineering or other corrective actions. Such actions could reverse the trend by the 24th month and produce a positive acceleration in job attitude as the reenlistment decision point is approached.

Ladders with profiles such as the Dental Technicians show high positive job interest with little relationship between interest and reenlistment. If reenlistment quotas are not being met in this ladder, the solution is probably in monetary or other such incentives. The ladder shows little potential return for effort made to improve the jobs.

Another use of the data base will be validation of statements of career intent with actual reenlistment decisions made. Extensive regression studies will be performed within and between career ladders to investigate the inter-relationships of job factors, background and aptitude factors, and job attitudes and career decisions.

As stated earlier, factors related to dissatisfaction reports are essentially unique to each career ladder. There are, however, a few relationships which have been consistently found. Those job incumbents whose jobs are more demanding and varied and are most related to their aptitude, education, and Air Force training are more satisfied with their jobs and tend to express positive predispositions toward Air Force careers. Regression studies indicate that job difficulty, number of tasks performed, and aptitude requirements of the job are positively related to job attitude while aptitude, education, and TAFMS for 1st-term airmen are negatively related.

Studying individual career ladders leads to specific restructuring of jobs and suggested changes in the classification system, technical training, and aptitude and educational requirements. However, the process places great demands on limited professional resources. General guidelines and methodology will assist management engineering and classification personnel in such actions. A potential source for operationally reviewing specific job structures for improvement has hardly been tapped; i.e., the job incumbents. With this in mind, a pilot study has sent occupational surveys to personnel in the Medical Specialist and Ground Radio Communication specialties. The surveys request each job incumbent to rate all tasks in his specialty according to his preference for performing the task. A second survey requested the Medical personnel to structure "ideal jobs." The surveys have been returned and will be subjected to cluster analyses to see if the incumbents can provide practical clusters of tasks that form interesting and meaningful jobs.

In summary, the Air Force is two years into its program of job satisfaction research. The goal is to sustain an all-volunteer force through making maximum use of available talent. The operational concern is to identify the dimensions of job satisfaction, to measure the satisfaction levels of these dimensions, to identify problem areas which have the greatest potential for improvement, to measure the complex interplay of specific job content, changes in job content, and characteristics of individuals on job attitudes, and finally, to reengineer jobs accordingly.

It is a big order, and the program has limited manpower resources. Notwithstanding, preliminary studies are well in progress. Current projects include development of an Occupational Attitude Inventory and an Assignment Preference Survey; establishing levels of job satisfaction for the airmen force in general and for specific career ladders; indepth

studies of specific career ladders evidencing high levels of job dissatisfaction to identify causes and suggest remedial actions; development and utilization of a massive data base to identify potential career ladders for detailed study and to provide a medium for between and within career ladder studies. Special studies are being conducted to identify critical supervisory experiences, to evaluate the ability of surveys to eliminate the need for field interviews, and to permit job incumbents to have input into structuring jobs.

The most crucial studies are in the design and planning stages -- longitudinal studies which will relate changes in job content to changes in job attitudes and eventual career decisions.

REFERENCES

Gould, R. B. Reported job interest and perceived utilization of talents and training by airmen in 97 career ladders. AFHRL-TR-72-7, AD-745 099. Lackland AFB, Tex.: Personnel Research Division, Air Force Human Resources Laboratory (AFSC), January 1972.

Gould, R. B. & Christal, R. E. Enlisted job satisfaction in the Air Force: A study at the task level. In Proceedings: 13th Annual Conference of the Military Testing Association. U. S. Marine Corps, Washington, D. C., 20-24 September 1971, Pp. 323-333.

Herzberg, F. Work and the nature of man. New York: The World Publishing Company, 1966.

Smith, P. C., Kendall, L. M., & Hulin, C. L. The measurement of satisfaction in work and retirement. Chicago: Rand-McNally, 1969.

Tuttle, T. C. & Hazel, J. T. Review and implications of job satisfaction and work motivation theories for Air Force research. AFHRL-TR-73-(In Press). Lackland AFB, Tex.: Occupational Research Division, Air Force Human Resources Laboratory (AFSC), 1973.

Weiss, D. J., Dawis, R. V., England, G. W., & Loftquist, L. H. Manual for the Minnesota satisfaction questionnaire. Minnesota Studies in Vocational Rehabilitation XXII, 1967.

Whitsett, D. A. & Winslow, E. N. An analysis of studies critical of the motivator-hygiene theory. Personnel Psychology, 1967, 20(4), 391-416.

A METHODOLOGY FOR DETERMINING JOB SATISFACTION IN THE U. S. NAVY

Lawrence A. Goldman

Navy Occupational Task Analysis Program, Washington, D. C.

In the armed services, in effect an All Volunteer Force, it is essential that job satisfaction be maximized while job dissatisfaction be minimized. There are three major reasons for doing so. First, satisfied individuals are more likely to make the military a career. Research conducted on personnel in private industry has indicated that the higher the overall job satisfaction, the lower the turnover rate. Because of substantially higher personnel costs needed to maintain an All Volunteer Force, it is important to keep the attrition rate as low as possible.

Second, knowledge of the sources of satisfaction and dissatisfaction will enable the military to more effectively utilize its limited available personnel resources. This is especially true regarding the composition of the personnel within any occupational career field which may be affected under the All Volunteer Force concept. For example, in an occupational career field in which men predominate, the number of women might substantially increase. In view of this possibility, it would be important to determine whether or not women are satisfied in this field, in order to help decrease their rate of attrition.

Third, an understanding of the sources of satisfaction and dissatisfaction makes it possible for the armed services to improve the general well-being of its personnel, a major concern in its own right.

The purpose of this study is to explain a methodology for identifying the major sources of satisfaction and dissatisfaction for four Navy ratings (occupational career fields for enlisted personnel) in the aviation community. The focus of this study, however, is upon the Aerographer's Mate (AG) rating, an occupational career field comprising personnel trained as weather forecasters. For the AG rating, this methodology is used also to examine the relative satisfaction between groups based upon military paygrade, sex, mental ability, and reenlistment intention. These four variables were selected because of their potentially important bearing on the success of an All Volunteer Force.

The data were gathered by use of a job satisfaction questionnaire contained within the task inventory booklet for each rating. Each individual was required to provide responses to statements associated with 33 job factors which emerged from an extensive study conducted by the Canadian Forces Personnel Applied Research Unit. For the AG rating, three methods of administration were used to collect the data from 727 Aerographer's Mates at 82 different activities in paygrades E-2 through E-9, representing 59% of the total strength of this rating: (a) Administration by teams of occupational analysts from the Navy Occupational

Task Analysis Program (NOTAP); (b) command proctoring by an individual designated by the commanding officer of the activity surveyed; and (c) individual mailout to selected individuals throughout the fleet. For subsequent data collection efforts, the individual mailout method was discontinued because of a poor rate of return.

For each of the 33 job factors, satisfaction was viewed as a discrepancy between responses to a "How much is there now?" item and a "How much should there be?" item. Using a five point Likert scale (ranging from minimum with a value of "1" to maximum with a value of "5"), the discrepancy for each participant was determined. Theoretically, an individual who perceived a particular job factor to be as available to him on the job (How much is there now?) as he would like it to be (How much should there be?) would rate these items the same. A higher rating on "How much should there be?" than "How much is there now?" indicated dissatisfaction. Specifically, the magnitude of the difference between the two responses on each job factor was viewed as a measure of the extent of the job factor's deficiency. An individual who responded "1" on the "How much is there now?" item and "1" on the "How much should there be?" item would be just as satisfied as someone who responded "5" and "5" respectively. An individual responding "1" and "5" respectively, though, would be very dissatisfied with that particular factor since the difference has the maximum magnitude of 4.

Difference or discrepancy scores were obtained for each individual on each job factor. In the analysis of the AG rating, these discrepancy scores were then used as input for single classification analysis of variance combined with a composite t test as described by Garrett in the 6th edition to his Statistics in Psychology and Education.

In the event that the analyst only has access to the number of individuals, means, and standard deviations, then the analyst can run t tests on each pair of job factors sequentially, the type of t test used depending upon whether or not there is homogeneity of variance. The major difference between these two approaches can be explained as follows. Only a single standard deviation, the within group or error standard deviation, is computed for the composite t test used with analysis of variance. This error standard deviation is based upon all scores and with generally high error degrees of freedom. On the other hand, as many standard deviations as factors (or variable categories) must be computed whenever separate t tests are run. According to Garrett, the within group standard deviation furnishes a better or more reliable measure of uncontrolled variation than could be obtained from separate standard deviations. The results obtained from both approaches are essentially the same, though, whenever the means are based upon a substantially large number of cases. Moreover, whenever the number of cases is large, it is noted that it essentially makes no difference whether or not there is homogeneity of variance. The other three ratings in this study, the Aircrew Survival Equipmentman (PR) rating, the Photographic Intelligenceman (PT) rating, and the Aviation Antisubmarine Warfare Operator (AW) rating, were examined by the use of separate t tests since individual case data were not available as they were for the AG rating.

Regardless of the type of t test used, statistically significant differences

($p < .05$) between job factors and between groups based on the selected variables were then determined. Results obtained from applying either approach were then used to place the job factors or the variable categories into separate clusters. Pertinent to these clusters, job factors or variable categories appeared in two clusters whenever they were statistically equivalent to two or more other factors (or categories) which were statistically different from each other.

Table 1, which is found on page 6, indicates the major sources of satisfaction and dissatisfaction for the AG rating. Ten clusters of these 33 job factors, together with the mean discrepancy for each factor, are noted in this table. In the AG rating, pay and allowances (cluster 1) was the most dissatisfying aspect of the job, while adequacy of living conditions, recognition for work done, and proper utilization of money and/or resources (cluster 2) were more satisfying than pay and allowances but more dissatisfying than the other factors. Clusters 3 through 10 can be similarly interpreted. It is observed that freedom from frequent job changes in the activity (cluster 10) was the most satisfying job factor, perhaps because Aerographer's Mates believed that they experienced job stability to a greater extent than generally exists in other Navy ratings.

Table 2, which is found on page 7, summarizes the differences which exist between military paygrades on 21 of the 33 job factors for the AG rating. This table indicates that whenever differences occurred between paygrades, generally E-6's and above were more satisfied than E-5's and below except for opportunity for promotion and adequate operation of the military system. A close examination of this table reveals that while E-8's and E-9's were the most satisfied individuals, E-5's were the most dissatisfied. In order of increasing dissatisfaction, E-7's, E-6's, E-2's, E-3's, and E-4's were more dissatisfied than E-8's and E-9's but more satisfied than E-5's.

Analysis of the job factors for the AG rating based on sex revealed that women were as satisfied as men except for pay and allowances in which women were more satisfied than men. Despite this one difference, men and women were equally satisfied since one or two significant differences could occur by chance based upon 33 separate tests.

Analysis of basic mental ability is based upon the combined General Classification Test and Arithmetic Test score of the Navy's Basic Test Battery, in which these six ranges of scores were used: (a) less than 100; (b) 100-109; (c) 110-119; (d) 120-129; (e) 130-139; and (f) 140 and higher. The composite General Classification Test and Arithmetic Test score was used because the normal minimum requirement for formal school ("A" school) training in the AG rating is a composite score of 110 or more. Review of mental ability in this rating indicated no significant differences between these six categories, except that the highly intelligent 130-139 group was more dissatisfied than the other mental ability groups on freedom from restrictive rules and regulations, worthwhileness of work, and opportunity for helping others. Even this exception is not conclusive since the highest mental ability group (140 and higher) was as satisfied as the other mental ability groups on these three job factors.

In addition to the background variables of military paygrade, sex, and mental ability, the reenlistment intentions of AG's were analyzed. Concerning their reenlistment intention, each individual responded to the question "Do You Plan to Reenlist?" by either "Yes", "No", or "Uncertain". As shown in Table 3, which is found on page 8, there were significant differences for 20 of the 33 job factors. On eight factors, individuals planning to reenlist and individuals uncertain about reenlistment were more satisfied than personnel not planning to reenlist. On ten factors, individuals planning to reenlist were more satisfied than personnel uncertain about reenlistment and personnel not planning to reenlist. On the other two factors, training for the job and guidance received on the job, individuals planning to reenlist were more dissatisfied than both personnel uncertain about reenlistment and personnel not planning to reenlist. Concerning these two factors for which those planning to reenlist tended to be more dissatisfied, perhaps additional job training and guidance should be provided AG's as inducement for a more satisfying career.

The major purpose of the methodology used in this study was to provide a better assessment of job satisfaction and dissatisfaction than could be achieved through rank ordering. Pertaining to Table 1, pay and allowances would rank highest if rank order were used (that is, a rank of 1 in terms of greatest dissatisfaction). Adequacy of living conditions would be assigned a rank of 2, recognition for work done would be assigned a rank of 3, etc. A rank ordering would imply that adequacy of living conditions was more dissatisfying than recognition for work done, etc. Based upon statistical criteria, however, there were usually clusters of two or more job factors that were equally satisfying (or dissatisfying) statistically.

An important aspect of this approach is that reliable results would be obtained regardless of how job satisfaction was defined. There are at least nine different operational definitions of job satisfaction which are cited in the literature, including: (a) The "Is Now" model, in which job satisfaction is defined in terms of responses only to a "How Much is There Now?" item for each job factor; and (b) models in which importance ratings are used to weight various combinations of "How Much is There Now?" and "How Much Should There Be?" items (for example, Importance X (Should Be - Is Now), Importance X Is Now, and Importance - Is Now).

This approach can be applied to any rating or combination of ratings. In particular, the results obtained from analysis of the AG rating can be compared with other naval ratings in order to determine the job factors which are generally indicative of greatest satisfaction and dissatisfaction in the Navy. As previously indicated, this has been done already for three other ratings - the Aircrew Survival Equipmentman (PR) rating, the Photographic Intelligenceman (PT) rating, and the Aviation Antisubmarine Warfare (AW) rating. The sources of greatest satisfaction and dissatisfaction for these three ratings (together with those for the AG rating) are shown in Tables 4 and 5 on pages 9 and 10 respectively. The factors which represent these major sources comprise roughly the highest and lowest 25% of the 33 job factors in each rating, based upon either

the first three clusters of factors or the first eight factors regardless of the number of clusters in which these factors appear. The frequency of occurrence of the 13 factors representing the sources of greatest satisfaction and the 13 factors representing the sources of greatest dissatisfaction in these four ratings is shown in Table 6 found on page 11. Review of Table 6 indicates that while individuals are given the opportunity to assume responsibility in carrying out their assigned duties, and while they get along well with their coworkers, they are dissatisfied with what they perceive to be sufficient recognition for their efforts as well as with their overall growth and development. Pay and allowances, which appears as a major source of dissatisfaction in these four ratings, is probably more satisfying at the present time in view of the recent pay increases. It is noted that opportunity for promotion appeared as a major source of dissatisfaction in three of these four ratings. A substantially greater number of ratings should be analyzed, however, before general trends can be discerned Navy-wide, or at least within the aviation community.

This methodology could also be applied to occupational career fields in the other branches of the armed services. Through the use of this methodology, recommendations could be made regarding the sources of job dissatisfaction that should be minimized in order to achieve a more effective military organization, especially under the All Volunteer Force.

TABLE 1

Relative Satisfaction of Aerographer's Mates on 33 Job Dimensions

Cluster	Job Dimensions	Mean Discrepancy
1	Pay and allowances	2.04
2	Adequacy of living conditions	1.87
	Recognition for work done	1.79
	Proper utilization of money and/or resources	1.75
3	Opportunity for promotion	1.53
	Opportunity for growing and developing	1.44
	Job appeal	1.43
	Adequate operation of the military system	1.42
	Opportunity for achievement	1.41
4	Opportunity for growing and developing	1.44
	Job appeal	1.43
	Adequate operation of the military system	1.42
	Opportunity for achievement	1.41
	Opportunity to have pride for the service	1.40
	Fair treatment by the system	1.40
	Adequacy of work surroundings	1.38
	Training for the job	1.31
5	Training for the job	1.31
	Competence of seniors	1.25
	Job challenge	1.23
	Resources to do the job	1.21
	Opportunity for proving self	1.20
	Status within the organization	1.18
6	Status within the organization	1.18
	Chance to do the primary job	1.06
7	Chance to do the primary job	1.06
	Competence of subordinates	1.05
	Acceptance of your recommendations	1.04
	Opportunity to do the entire job	1.02
	Worthwhileness of work	1.00
	Adequacy of relations with superiors	0.98
	Guidance received on the job	0.97
	Opportunity to see work results	0.95
	Opportunity to contribute	0.95
	Freedom from restrictive rules and regulations	0.94
8	Opportunity to see work results	0.95
	Opportunity to contribute	0.95
	Freedom from restrictive rules and regulations	0.94
	Opportunity for helping others	0.87
	Responsibility and independence in carrying out the job	0.84
9	Opportunity for helping others	0.87
	Responsibility and independence in carrying out the job	0.84
	Freedom from job pressures	0.78
	Adequacy of relations with subordinates	0.76
10	Freedom from frequent job changes in the activity	0.56

TABLE 2

Job Dimensions for which Differences Exist between Paygrades

Job Dimensions	Differences which Exist
Pay and allowances Opportunity for growing and developing Opportunity for proving self Chance to do the primary job Opportunity to do the entire job Worthwhileness of work Adequacy of relations with superiors Opportunity to contribute Opportunity for helping others Adequacy of relations with subordinates Fair treatment by the system	E-6's and above were more satisfied than E-5's and below
Recognition for work done Opprtunity to see work results	E-8's and E-9's were more satisfied than E-7's and below
Opportunity for promotion Adequate operation of the military system	E-4's and below were more satisfied than E-5's and above
Proper utilization of money and/or resources Opportunity for achievement	E-7's and above were more satisfied than E-6's and below, with E-5's being more dissatisfied than the other paygrades
Job challenge	E-6's and above were more satisfied than E-5's and below, with E-2's and E-3's being more satisfied than E-4's and E-5's
Adequacy of relations with superiors	E-4's and E-5's were more dissatisfied than the other paygrades
Freedom from restrictive rules and regulations	E-5's and above were more satisfied than E-4's and below
Responsibility and independence in carrying out the job	E-6's, E-8's, and E-9's were more satisfied than the other paygrades

TABLE 3

Job Dimensions for which Differences Exist between Reenlistment Intentions

Job Dimensions	Differences which Exist
Job appeal Responsibility and independence in carrying out the job Opportunity for promotion Opportunity for proving self Worthwhileness of work Job challenge Fair treatment by the system Proper utilization of money and/or resources	Individuals planning to reenlist and individuals uncertain about reenlistment were more satisfied than personnel not planning to reenlist
Recognition for work done Freedom from restrictive rules and regulations Freedom from job pressures Competence of seniors Adequacy of relations with superiors Opportunity for growing and developing Opportunity to have pride for the service Opportunity to contribute Opportunity to do the entire job Opportunity for helping others	Individuals planning to reenlist were more satisfied than personnel uncertain about reenlistment and personnel not planning to reenlist
Training for the job Guidance received on the job	Individuals planning to reenlist were more dissatisfied than personnel uncertain about reenlistment and personnel not planning to reenlist

TABLE 4

The Sources of Greatest Job Satisfaction in the AG, PR, PT, and AW Ratings

AG Rating	PR Rating
Opportunity to see work results Opportunity to contribute Freedom from restrictive rules and regulations Opportunity for helping others Responsibility and independence in carrying out the job	Opportunity to contribute Adequacy of relations with superiors Worthwhileness of work Adequacy of relations with superiors Worthwhileness of work Opportunity to see work results
Opportunity for helping others Responsibility and independence in carrying out the job Freedom from job pressures Adequacy of relations with subordinates Freedom from frequent job changes in the activity	Opportunity to see work results Opportunity for helping others Opportunity for helping others Adequacy of relations with subordinates Responsibility and independence in carrying out the job

PT Rating	AW Rating
Competence of subordinates Opportunity to see work results Freedom from job pressures Opportunity to do the entire job Adequacy of relations with superiors Freedom from frequent job changes in the activity Responsibility and independence in carrying out the job	Job challenge Opportunity to contribute Adequacy of relations with superiors Responsibility and independence in carrying out the job Opportunity for helping others Freedom from frequent job changes in the activity Adequacy of relations with subordinates
Adequacy of relations with superiors Freedom from frequent job changes in the activity Responsibility and independence in carrying out the job Adequacy of relations with subordinates	Freedom from job pressures

TABLE 5

The Sources of Greatest Job Dissatisfaction in the AG, PR, PT, and AW Ratings

AG Rating	PR Rating
Pay and allowances	Pay and allowances
Adequacy of living conditions	Adequacy of living conditions
Recognition for work done	Opportunity for growing and developing
Opportunity for promotion	Job appeal
Opportunity for growing and developing	Adequate operation of the military system
Job appeal	Opportunity for achievement
Adequate operation of the military system	Recognition for work done
Opportunity for achievement	Opportunity for promotion
	Proper utilization of money and/or resources

PT Rating	AW Rating
Proper utilization of money and/or resources	Proper utilization of money and/or resources
Pay and allowances	Adequacy of living conditions
Recognition for work done	Recognition for work done
Adequacy of living conditions	Adequate operation of the military system
Opportunity for growing and developing	Opportunity for promotion
Adequate operation of the military system	Opportunity for promotion
Status within the organization	Adequacy of work surroundings
Training for the job	Pay and allowances
	Fair treatment by the system
	Job appeal

TABLE 6

Frequency of Occurrence of the Sources of Greatest Job Satisfaction and Dissatisfaction in the AG, PR, PT, and AW Ratings

Sources of Greatest Job Satisfaction	Frequency of Occurrence
Responsibility and independence in carrying out the job	4
Adequacy of relations with subordinates	4
Opportunity to see work results	3
Opportunity to contribute	3
Opportunity for helping others	3
Freedom from job pressures	3
Freedom from frequent job changes in the activity	3
Adequacy of relations with superiors	3
Freedom from restrictive rules and regulations	1
Worthwhileness of work	1
Competence of subordinates	1
Opportunity to do the entire job	1
Job challenge	1

Sources of Greatest Job Dissatisfaction	Frequency of Occurrence
Pay and allowances	4
Adequacy of living conditions	4
Recognition for work done	4
Adequate operation of the military system	4
Opportunity for promotion	3
Opportunity for growing and developing	3
Job appeal	3
Proper utilization of money and/or resources	3
Opportunity for achievement	2
Status within the organization	1
Training for the job	1
Adequacy of work surroundings	1
Fair treatment by the system	1

REFERENCES

Dawis, R., & Weitzel, W. The measurement of employee attitudes. Minneapolis: The Center for the Study of Organizational Performance and Human Performance, University of Minnesota, 1971.

Garrett, H. E. Statistics in psychology and education. (6th ed.) New York: David McKay Company, Inc., 1966.

Goldman, L. A. Occupational analysis: Report of analysis of job satisfaction data for the Aerographer's Mate rating. WTR 73-13. Washington, D. C.: Naval Personnel Research and Development Laboratory, 1973.

Lawler, E. G. Job attitudes and employee motivation: Theory, research, and practice. Personnel Psychology, 1970, 23, 233-237.

Otke, P. G. The relationship between job satisfaction and job requirements. Technical Report 70-2. Toronto: Canadian Forces Personnel Applied Research Unit, 1970.

Porter, L. W. A study of perceived need satisfactions in bottom and middle management jobs. Journal of Applied Psychology, 1961, 45, 1-10.

Vroom, V. H. Work and motivation. New York: John Wiley & Sons, 1966.

Walters, L. K., & Roach, D. Relationship between job attitudes and two forms of withdrawal from the work situation. Journal of Applied Psychology, 1971, 55, 92-94.

Wanous, J. P., & Lawler, E. E. Measurement and meaning of job satisfaction. Journal of Applied Psychology, 1972, 56, 95-105.

INTEREST SURVEYS AND JOB PLACEMENT IN THE ARMED SERVICES

William J. Figel
Science Research Associates, Inc.
259 East Erie Street
Chicago, Illinois 60611

Paper presented at
15th Annual Conference of The
Military Testing Association
Oct. 28 - Nov. 2, 1973
San Antonio

With the demise of the draft, the armed services no longer have a captive pool from which to create their future soldiers, sailors, and airmen. Instead, they must attract and retain recruits in competition with business and industry. In addition, through the "up or out" policy, the armed services are trying to upgrade significantly the quality of their work force, thus putting themselves even more directly in competition with industry for the ablest workers.

The military seem to be hampered in this competition by a number of restraints--their on-the-surface wage scales are inflexible and generally inferior to those offered by industry, their recruits cannot often be employed near their own homes, their training is more rigorous and regulations more restrictive than those commonly encountered in industry. The armed services have already initiated significant changes in all these areas, but none that seem likely to overcome completely the competitive disadvantage. In fact, present levels of recruitment are notably short of what is needed, and the proportion of Class IV recruits is dangerously high. The problem from the point of view of the recruiting officer is almost one of a textbook case in personnel management: how can a firm in this situation maintain an adequate work force and avoid excessive turnover?

The textbook solution is easy to state as a general principle --it is to emphasize, and even create, areas of competitive advantage. Significant efforts along these lines are also in evidence. Billboards proclaim that military life is more interesting, more challenging, than other ways of life. Enlistment is heralded as the opportunity to learn a trade. Even the necessity to live away from home may be seen as an inducement, rather than a disadvantage. The purpose of these remarks is to consider interest surveys within the context of a job placement program that may offer another competitive advantage over the alternatives offered to the young adult by business and educational institutions.

There is no doubt that money in the pocket provides a strong motivation for a workman to seek a new job or to stick to his present task. But it is not the only motivation. Entrepreneurs who are ready to turn their hands to whatever activities offer the largest profits are presumably still to be found on the staffs of large conglomerates, and they are undoubtedly the sine qua non of crime syndicates, but they are by no means typical of the young adults who are the targets of present military recruiting efforts. A pertinent study has just recently been reported in the Vocational Guidance Quarterly (1). A sample of Army and Air Force personnel stationed in West Berlin were asked to rate a number of personal and economic

needs in terms of their importance to the individual's satisfaction with his job, and also in terms of the degree to which these needs were satisfied by the individual's current military activities. The results of the study suggested that needs for security and socio-economic status were being met to an even greater extent than they were felt to be needed, while needs related to self-actualization--such as self-expression and creativity--were not being met to the extent they were felt to be needed. In short, these men tended to feel, though they might never have replied so to a direct question, that they did not need more money as much as they needed a task better in accord with their own sense of themselves. This particular study encompassed personnel of all grades from junior enlisted personnel to senior officers. There are many indications that among young people the same feeling is even more prominent. Today's youth are seeking more than a high salary. Of equal or greater importance to them is a match between their interests and the challenge presented by the job.

Let's assume for the moment that there are, somewhere in the nation, a goodly number of young people to whom life in the modern armed services would not be largely uncongenial. Let's suppose that most of them are mature enough to know that no job anywhere will satisfy their wildest dreams every moment of the day. And let's suppose further that some of them present themselves to a recruiting officer. Possibly at present he has no task more urgent than to enlist them as quickly as possible.

But one key basis for a volunteer army is not just filling available positions, but filling them with applicants who can do the job well; and not only applicants who can do the job well, but applicants who are likely to want to continue to do the job when their initial enlistment period runs out. I am sure it is no news to members of the Military Testing Association that identifying such applicants means taking interests into account. In fact, in my preparatory review of the literature on this subject I found that the ground had been well prepared by Allen Schuh, of the Naval Aerospace Medical Institute in Pensacola. As a result of a thorough survey of studies published up to 1967 (2), he concluded that interest inventories are better predictors of employee tenure than are either intelligence or aptitude tests. Not much has turned up since then to challenge that conclusion.

Of course, this is not meant to disparage the use of aptitude or ability tests. The typical pattern in industrial employment, not surprisingly, is that groups of employees whose interest and abilities both match their jobs will have the longest average tenure, while groups of employees who have a mismatch in either interests or abilities will have about the same but shorter average tenures. And employees who don't match the job in either interest or ability don't last at all, (except possibly in the army during wartime).

Neither is it any secret that taking interests into account means some kind of systematic interest survey. It is not enough to ask a young man whether or not he would like to be a movie actor or a hogshead hooper. In the one case he may be thinking of John Wayne while you are thinking of a wound victim in a training film; in the other case he may have no idea at all of what is meant. But to show that I am well aware I am carrying coals to Newcastle, I might mention that the subject was well covered in 1963 in A Guide to Psychological Tests, by Joseph Zeidner, who was then chief of the Support Systems Research Laboratory of the U.S. Army Personnel Office. As he put it, "it has been well demonstrated that interests cannot always be validly measured by direct questions, especially by people with little experience or relevant information." (3)

I tend to disagree with Zeidner's review only about his preference for manifest and tested measures of interest. A person's manifest interests are determined by things he had done--he is interested in carpentry because he has built a boat, or in driving because he has a driver's license. A person's tested interests are measured by his acquaintance with the technical vocabulary of various fields--if he thinks a bit belongs in a horse's mouth, you make him a cavalryman and not a computer programmer. But the same problems arise with these approaches as with direct questions: few of today's new recruits will have the opportunity to gain experience or under-

standing in some of the most critical areas of employment in the modern armed services, such as radar technology or computer programming, for example.

The alternative to looking at manifest or tested interests is to use an interest inventory, such as the Kuder Occupational Interest Survey or the Strong Vocational Interest Blank. These are probably the two most widely used of such instruments, and since SRA has a vested interest in one of them, the Kuder survey, we might take a moment to describe some of their differences. Strong's questionnaire asks subjects to rate various activities on a like-dislike scale, while Kuder's asks them to pick the most liked and least liked of a set of three activities. Form C of the Kuder produces profiles on a set of ten normed scales with titles like Outdoor, Computational, Persuasive, Clerical, etc. The entire profile can be compared with average profiles for many different occupations. Strong's instrument provides 22 scales, with titles directly associated with occupational groupings, as well as separate scales for different occupations. Kuder's form DD gives a numerical index for the degree of match between the subject's responses and the average responses of a criterion group in different occupations.

But the general nature of these instruments is undoubtedly well known to anyone much involved in military testing. A recent listing of our industrial and government customers for

Kuder C and Kuder DD shows a Veteran's Administration Hospital on nearly every other page, several military rehabilitation programs, and even some Army, Navy, and Air Force bases. In fact, the widespread use of interest surveys in a program of careful job placement may constitute a competitive advantage for the armed services in attracting and retaining personnel. We know there are many commercial firms that use these surveys, but we also know there are a great many more that do not. We tend to suspect that the greatest use of interest surveys is made in placement at the management level, and we fear that hourly workers are almost forgotten. There is probably a tendency in industry to hire whomever they can get during periods of manpower shortage, and to hire only the most able during periods of manpower surplus--with the idea, perhaps, that the new employee is darn lucky to get the job, and will remain everlastingly grateful. It isn't always so. In fact, you of the Military Testing Association might take pride in your role in inspiring the following bit of advice, in a text on personnel management: "A young 18-to-20-year-old high school or technical school graduate will be the skilled machinist, experienced planner, valued draftsman, or even foreman of the 1970's and 1980's. He needs as careful selection as the college graduate, but he rarely gets it from industry. That is one reason why many of the brighter youngsters with potentials in these areas are looking at other careers--primarily in the

armed services and in government." (4) If this does not entirely square with your experience, perhaps some new factors have entered the situation since 1967, when this was written.

Nevertheless, the educational and training opportunities in the armed services are exceptional, and are rightly advertised as such. No industrial concern offers such a broad range of careers, and no educational institution has such an excellent program of job placement for its trainees. I think it is clear that interest surveys can well serve the interests of recruits to such a program.

But of course we have also the interests of the services themselves in mind. There are peculiarities to interest surveys that allow them to be used for purposes for which they have very little face validity. The Kuder Form DD, for instance, asks for no choices which would seem to distinguish between electrical engineers and heating and air-conditioning engineers, yet in the process of averaging patterns of responses for groups of people in these two professions, we come up with scales that do a pretty fair job of distinguishing between the two groups. We do a better job distinguishing between electrical engineers and lawyers, to be sure, but some apparently close distinctions can still be usefully made. We have had some success in predicting academic performance

from interest survey responses alone. And there are several studies of their usefulness in identifying psychological abnormalities.

All of this is by way of suggesting the usefulness of interest surveys in identifying recruits who are not only able to do the various tasks that the modern armed services need to have done; but who will be satisfied to do them under the conditions of life that the armed services must impose. Provided that there are computer programming specialists who re-enlist even though they might find more highly paid jobs elsewhere, some experience with their interest patterns is likely to be invaluable in selecting the new recruits "most likely to succeed" in terms of successful job performance over a suitably long period of time.

(1) Huskey, Hyrum H., Jr., "Expressed Needs and Jobs Satisfaction among Military Personnel," The Vocational Guidance Quarterly v. 22, no. 1; Sept. 1973

(2) Schuh, Allen J., "The Predictability of Employee Tenure: A Review of the Literature", Personnel Psychology, v. 20 (1967) pp. 133-152

(3) Zeidner, Joseph, Guide to Psychological Tests, Prentice-Hall, New Jersey, 1963; (Ch. 4)

(4) Morgan, John S., Managing the Young Adults, American Management Association, Inc., 1967 (p. 46)

Job Description in the Armed Forces as a Basis for Analysis and Evaluation

Flottillenadmiral Fiebig
German Navy

Opinions, as we all know, differ, sometimes widely, on the possibilities and value of describing, analysing and evaluating jobs and duty positions.

In the Federal Republic of Germany the legal situation, under the "Act on the Constitution of Enterprises" since 1972 had been that private employers are required to provide employees with concrete information concerning positions, in particular with respect to

- the nature of the job,
- its place in the organization,
- the duties and responsibilities associated with it.

Although the Act does not elaborate on this point, one of its effects is bound to be some measure of systemization of job descriptions in trade and industry. First beginnings in that direction are already being made.

Thus, the Federal Ministry of Labor and Social Affairs, for instance, promoted and helped to establish an institution called "Committee on Man in the Administration" whose aim it is to promote the cooperation of experts from the scientific community, industry, the Public Service, the Trade Unions, Employer Associations, and the Labor Administration so that common solutions to the many problems of quantitative and qualitative manpower planning, development and administration might be developed and propagated, using modern scientific findings and technologies. In addition, a number of private organizations, such as the "Academy for Work Science" and the "HAY Business Consultants" firm, for instance, are engaged in research, teaching, and consultative work in the field of structural investigations, work analysis, and job description. Other organizations, such as the German Shell AG, for instance, have developed their own techniques and procedures.

Independent thereof, an analogous trend for the entire public service sector of the Federal Republic has become unmistakable since, in 1973, a commission appointed by the Federal Government to study the reform of public service law submitted concrete proposals on how the public career and pay system should be objectivated by putting it on a function-oriented footing. One of the measures recommended to that end was to develop a number of instruments which should first be tested in selected areas.

This is the larger framework in which the efforts and work of the Federal Ministry of Defense must be seen on which I will now report.

In the spring of this year (1973) we began with the testing of a set of instruments for surveying and reorganizing the jobs in the German Armed Forces. In other words, almost two years went by before the central, basic proposal contained in the report by the Manpower Structure Commission appointed by the Ministry of Defense - namely to develop an Armed Forces - specific system of job description, job analysis and job evaluation - had finally been realized to the point where we had an instrument ready to be tested.

In its investigations the Manpower Structure Commission found that the present evaluation and pay system in the Federal Armed Forces is, in the final analysis, based on rank. Rank, however, it determined, is no longer indicative of the duties incumbent on the individual holding such rank, for one and the same rank today often embraces a number of highly divergent jobs. Concrete individual jobs are hence subsumed under an abstract rank which in turn determines the rate of pay.

Past methods of job evaluation are neither up to modern standards nor do they fit the specific needs of modern armed forces. The content of a position, i.e., the job and the training and qualifications required to fill it, are not sufficiently taken into account, considering modern work science standards.

Relatively large differences in the degree of strain involved in otherwise identical jobs with a like rating, for instance, have a considerable disturbing effect. These differences are completely ignored in the current evaluation and pay system, which, again, has ill effects on personnel. Where unequal jobs are subsumed under the same rank, the differences and disturbing effects are even greater, of course.

Career progression, in the sense that an individual may climb on the pay scale, is possible in the Bundeswehr today almost only by way of promotion to a higher rank. The consequence is a large number of relatively high ranks without a correspondingly broad base of entrance grades. An individual may be promoted to a higher rank although the job he performs does not warrant a higher rating. Assignment planning and the evaluation and pay system in the Bundeswehr are insofar confronted with serious difficulties which are considered to be surmountable only by instituting a completely new system. In this endeavor it would appear expedient to apply to the majority of the jobs in the Armed Forces the proven survey and evaluation techniques used in the private sector. This would offer the advantage that the public and the private sector would move closer together in such fields as training, job equivalency and the desired interchange of personnel, as not only the evaluation techniques but also the occupational classification system evolved for industry could be drawn upon in revamping the

system of duty positions in the Armed Forces. A grouping-together of military jobs by more functional characteristics, which might become necessary in the process, could result in more efficiency. We will have to examine very carefully though which jobs are so peculiarly military that their evaluation requires a departure from generally applicable criteria.

The Manpower Structure Commission was guided by the basic idea underlying the present system, namely that public service law was originally concerned with discrete jobs. The ever increasing number of discrete jobs, however, today not only requires that they be systematically re-inventoried. They must also be ordered in larger groups and weighted by the degree of executing and directing responsibility they involve in order that the source of public service law may again become transparent and manageable in its further evolution. To that end, the Commission held, a system of job description, job analysis, and job evaluation must be developed which permits a differentiated assessment to be made of the individual jobs in their various dimensions. Based on such a system a new personnel structure for the Bundeswehr must be devised which satisfies the needs of modern armed forces. Let me summarize these needs, as we see them, in five main points:

1. All individual jobs required to enable the Armed Forces to fulfill their mission must be capable of being functionally ordered in specialty groups in a hierarchically structured organization which can be managed by modern methods of command and control.

2. The new personnel structure must be adaptable to all conceivable forms of war and to new weapon systems.

3. It must provide a flexible framework for the modeling of careers and of the pay system.

4. It must be such that training objectives and the nature and scale of measures necessary to advance the soldier in his career can be rationally derived from it.

5. It is to provide an essential precondition for the optimal integration of the armed forces in society.

This is the overall objective for a new personnel structure of the armed forces from which the research work had to start out. Here, our first concern was to develop a set of survey instruments that would satisfy two requirements: First, it had to be capable of collecting, in surveys to be conducted in the armed forces, all the data needed for making structural improvements within the framework of the objectives mentioned. Second, it had to be such that it could be administered by the Bundeswehr itself. This, we considered, was a sine qua non, as otherwise the costs would be prohibitive.

The Federal Ministry of Defense commissioned a scientific institute to develop a system for job description, job analysis and job evaluation in the Bundeswehr.

As you know, there have been developed in Germany and in other countries quite a number of methods and technqiues for job analysis and job description. Having examined them, however, we felt that none of the procedures could be adopted without modification. Most of the procedures had been tailor-made to fit the specific needs of the industrial organization concerned and are not suited for the armed forces for several reasons.

To mention first some organizational and quantitative aspects: For one, these enterprises are generally much smaller in terms of the number of personnel they employ than the armed forces of the Bundeswehr with their 500,000 servicemen. For another, in no other enterprise is the diversity of jobs anywhere as great as in the armed forces. And finally, the enterprise-tailored procedures use conventional methods of job analysis and job description. Such an approach, however, is not feasible for the armed forces if only because of the large personnel effort that it would involve.

Apart from these largely organizational and quantitative aspects there is of course the peculiarity of the duties and tasks in the armed forces to be considered which although they have many parallels with comparable jobs in the civil service and in industry, are nevertheless distinct from them and characterized primarily by the following quality: Soldiers must acquire skills and knowledge and possess and develop abilities which are to be applied only in a defense emergency.

The evolvement of an instrument for job description and job analysis was geared to one central objective: to create the precondition for a system of specialty groups under which all positions can be subsumed. This means, above all, that answers would have to be found to the following four questions:

1. How is the position to be rank-ordered?

2. What training does the position require?

3. What qualifications and abilities does the position require?

4. What other employment potential, vertical and lateral, does the incumbent of a position offer?

In the job description it is first stated what duties the incumbent has to perform. Such a breakdown of positions into duties will later on permit us to move from a distribution of duties "as is" to a distribution of duties "as should be". The duties themselves would hardly be changed, only their specific content and combination in a job. This concept of duties "as should be" is also reflected in the designation chosen in the Bundeswher for the instrument, namely "function analysis". The term denotes that

what is to be accomplished is not an analysis of tasks as actually performed but the collection and/or development, in the form of duties, of system-related functions as a basis for reorganizing the personnel structure.

This function-related description is hence oriented not so much on what an incumbent actually does, than on the result to be achieved in the form of duties, and on the conditions under which this must be accomplished.

To summarize:

Job description in the Bundeswehr is to serve four primary objectives:

1. To enable positions to be groups into standardized jobs, and these in turn into specialty groups.

2. To enable requirements profiles to be developed and thus a requirements-related selection of personnel for, and matching of personnel with, specific jobs to be accomplished.

3. It is to provide sufficient data for a requirements-oriented training concept to be determined.

4. It is to furnish the basis for making decisions on the personnel structure in organizational units.

Here the problem was to find a job element suitable to meet the desired objectives of reorganization. By "job" we in the Bundeswehr understand the aggregate of the duties performed in a position.

As "duties" we designate the smallest organizational elements of a job.

Therefore, the instrument was so designed as to allow the positions to be broken down into duties. It was important that the different contents and typical combinations of duties were inventoried for the ultimate overall survey. In terms of the scope of the undertaking alone this was an extraordinary -let me say quantitative - problem. The scope of the problem is evident from the following data:

The personnel classification system of the Bundeswehr comprises some 1,400 specialties. These are arranged in 110 specialty groups and these in turn in 23 specialty are "skill level", "shredout", "occupation" and, finally, "position".

The "skill level" indicates the level and range of knowledge and abilities in a specialty.

The "shredout" identifies part of a specialty.

"Occupation" designates the combination of a specialty with a skill level and possibly a shredout. The "position" finally identifies the job in the organizational structure.

In 1971 and 1972 a number of preliminary surveys were conducted. They served several purposes. Accordingly, different survey techniques were used. The objective of the first survey phase was to make an inventory of the various job elements, i.e., of the duties, tasks, knowledge and skills that make up military jobs.

First the method of open questioning was used. The respondents had to complete a three-part questionnaire. In part 1 they delineated their duties as performers; in part 2 they described the various tasks required to be performed in execution of the respective duties; and in part 3 they were asked to list job factors indicative of the training required and the type and degree of strain involved in their jobs.

The survey was conducted by specially trained teams of interviewers. Even during the first stage of the survey it proved necessary to interview as many as close to 1,800 incumbents from 40 different units of the three Services.

Next followed a methods test. In it the first draft of a procedure for surveying duties and tasks, especially with regard to allocating "basic tasks" to duties, was tested. At this point let me give you a brief definition of the terms "duty", "basic task", and "task" as we use them in the Bundeswehr. "Duty" we call that component of a job which can be derived from the mission and which provides the basis for determining the distribution of the responsibilities associated with duty positions and hence the basis for ranking them in the organizational structure. From the analysis of duties typological characteristics may be derived for the construction of requirements and qualification profiles. By "basic task" we understand that component of a job which is indispensable for the execution of a duty in a position and which provides the basis for the organization of training. The analysis of basic tasks yields occupational/technical characteristics for the construction of requirements and qualification profiles. As "tasks" we designate job elements which describe the full range of work activities in a job, regardless of whether these elements are particularly vital for the execution of the duty or duties or not.

Before I continue let me say that at this point of the development work we found ourselves in a blind alley, so to speak, for despite intensive efforts we did not succeed in finding an analytical tool with which we could have filtered out that certain central component meeting the criteria of a "basic task" as defined above. At this point the development work succumbed to what might be called an "atomization process" which resulted from the sheer overwhelming flood of "ever smaller" job elements. The necessary "reduction work" which therefore was required for methodological reasons had to ensure that now at least all those tasks were ascertained

which were similarly indicative of the requirements for occupational/technical training and the occupational/technical characteristics to be contained in requirements and qualification profiles as the vainly sought "basic tasks".

At the same time the methods test in this phase of the work also pertained to various scales used in rating the time spent on the various duties and tasks and the degrees of difficulty and strain involved.

On the basis of the data obtained during the first survey phase catalogues of duties and tasks were compiled for four selected and circumscribed areas and submitted to the incumbents sampled in the form of questionnaires. The incumbents first had to checkmark all duties which they performed in their positions. In addition, they had to rate the average time spent in performing a duty. Here, too, methodological difficulties were encountered since a mere percentage rating cannot do justice to the many specific duties which may occur only in exercises or are the object of training in instruction courses.

In further steps the respondents had to check all the tasks they performed in their position and subsume them under the various duties, as well as rate on a scale the frequency with which a task occurred in a duty. For methodological reasons, this time the questionnaires were administered both to incumbents as performers and to their immediate supervisors. The supervisors had to complete the questionnaire not for the duties and tasks of their own positions but for the duties and tasks of their subordinates, i.e., of the same individuals who had previously been questioned about their jobs as performers.

This methods test was administered to 240 performers and supervisors of selected positions from four areas. In the next survey phase the inventory of duties, tasks and other job factors from the first stage was combined with the test-improved part of the survey instrument. This led to the development of a catalogue of preliminary functions. Here the functions served as coarse identifiers of duties. They represented "supra-duties" so to speak, under which the duties formulated by the incumbents had to be subsumed.

These function questionnaires contained both partly standardized as well as open questions. In addition to subsuming duties under a function the incumbent had to state whether he performed the duty himself or supervised its performance. The function questionnaires pertained only to the self-performance of duties. If an incumbent supervised the performance of a duty in whole or in part he had to describe the work of supervision on a separate questionnaire. The surveys in this second stage covered some 2,900 incumbents from 65 different units in the three Services.

From the experience and analyses in phases 1 and 2 resulted an instrument which by furnishing a differentiating description of duties and other job elements was to solve the problem of an Armed Forces-wide uniform job description such that all relevant criteria were covered. The purpose, in addition to completing the function questionnaire, was to test the practicability of the instrument in terms of the time, mental and physical effort that respondents could reasonably be expected to expend on answering the questions. In accordance with the state of progress of the development work, all questionnaires this time contained mostly performulated answers so that the work of the respondents was largely reduced to checkmarking. In this methods test some 180 incumbents of selected positions were questioned.

In a final survey phase, the purpose of which was to fill in gaps left during survey phases 1 and 2 and thus to complete the systematic survey of job elements, 900 incumbents in some 20 schools of the Bundeswehr were questioned.

This stage proved necessary for two reasons: In reviewing each of the preceding surveys it had been found that the selected samples had never been fully covered, be it that "stand-ins" with other than the desired job characteristics were encountered, or be it that the desired number of incumbents of certain positions could not be reached. These "gaps" had to be filled in. But beyond that - and this was another consideration - our purpose in extending this survey to schools was to obtain, through this vocationally and technically especially qualifed group of individuals, an indirect validation of the completeness of previously ascertained duties and other job elements.

The result of the preliminary surveys to develop an instrument for job description can be summarized as follows: From the questioning of 4,600 incumbents resulted a collection of 65,000 data on duties and 280,000 data on tasks.

On average, each incumbent listed between 12 and 20 duties and between 60 and 80 tasks.

These data on duties and tasks had to be capable of being compared, selected, and arranged later on. For that purpose they were, for each position separately and together with the position's other characteristics, coded and/or textually documented and stored on magnetic tape. The original data from the preliminary surveys are of considerable importance inasmuch as they reflect the language of the performers of the duties themselves. Therefore, it is advisable that this language should not be overly edited in the reviews. The more distinctly the language of the performers is preserved in the subsequent statements of duties and tasks, the greater will be the reliability of the results.

Analysis of this great quantity of survey data is accomplished by means of programs developed for that purpose, which permit the data to be assembled and sorted as desired. A detailed categorical coding system which can be used not only for the data from the surveys but also for

other source material now enables duty catalogues to be compiled for certain specialty groups as an instrument for job description and job analysis.

After the preliminary surveys, the job description instrument was designed; it consists of two parts.

The first part "Instrument Part A" is composed of catalogues of questionnaires for the various specialty groups to question incumbents as performers. Each of these catalogues of questionnaires is organized into four sections with the following content:

Section 1 is a questionnaire to ascertain personal background data (of the incumbent) and job satisfaction. This section is the same in each catalogue of questionnaires.

Section 2 constitutes the main part. It contains questionnaires on the duties which are performed by the incumbent himself and/or whose performance he supervises. The scope of section 2 is of necessity variable. It is determined by the number of duties in the respective specialty group. The duties in section 2 are classified into various duty groups, namely:

- Duties which are typical of the respective area, i.e. specialty-group-related;

- general administrative duties, i.e., duties, which are not restructed to a certain area;

- supervisory duties, i.e., duties of incumbents with supervisory functions which occur more or less in all areas;

- atypical or extraneous duties; these are considered to be duties which properly belong in other areas but which are also performed by incumbents in the respective area (for instance voice radio communications in the field, hence a duty properly belonging in the area "communications", which, however, must be performed by the members of an armored infantry squad since in this subarea no position "voice radio communication" exists.

Section 3 contains questionnaires on vocational/technical training and further training which is conducted within the scope of the job.

Section 4 finally contains questionnaires on direct technical supervision of personnel in the performance of their duties.

In addition, each catalogue of questionnaires comes with detailed instructions regarding the answering technique. The answers to the questions are not entered in the questionnaires themselves but are given on separate response sheets and continuation sheets. To complete the response sheets respondents checkmark preformulated answers. The continuation sheets are provided for incumbents to write in information.

Since the administration of this instrument by the Bundeswehr itself is not so easy and since also questions of data analysis had to be verified, the testing of the instrument of job description was conducted in the following four specialty areas this year:

- infantry duties

- aircrew duties

- navigation

- personnel administration.

I will now show you on some viewgraphs sections of "Instrument Part A":

(Viewgraph 1)

These questions pertain to the personal background of the incumbent and his satisfaction with his job. Specifically, the items covered are:

1. the highest level of formal education attained

2. the service status of the individual

3. the type of career

4. length of service in present duty position.

And: (Viewgraph 2)

5. Present rank.

Then follow a few questions concerning the incumbent's attitude toward his position, such as:

6. the interest he takes in his present job;

7. his contentment with his pay, considering his training and qualifications;

8. whether he can apply his abilities in his job;

9. his contentment with his pay, considering the strain involved and the performance required;

10. comparison with other jobs;

11. comparison with similar jobs in respect of pay.

(Viewgraph 3)

This viewgraph shows a duty from the area of "infantry duties". The questions pertain to:

- the various tasks to be performed in the duty;

- the circumstances under which these tasks must be performed (difficulty);

- the frequency of occurrence, and

- the importance of the duty.

The answers are entered in a separate response sheet which I will show you in the next viewgraph.

(Viewgraph 4)

This is the sheet in which the answers to the questions contained in the catalogue of questionnaires are entered. The number of the respective question is marked on the response sheet by making an oblique stroke across the corresponding number on the response sheet. Only in 3 cases must the respondents write figures, all other notations are made by placing an oblique stroke in the appropriate box. This facilitates the work of the respondents and reduces the number of illegible entries.

(Viewgraph 5)

These questions pertain to the vocational/technical training/further training which the incumbent conducts in the units. The purpose of the questions is to establish the degree of strain involved in the position. Specifically, the questions pertain to:

- the number of individuals trained;

- the importance of the training (whether the objective is earning a certificate of qualification, a permit, a license, an authorization, certification of a certain level of proficiency);

- on-the-job training;

- training through academic instruction (frequency);

- ascertaining the training result;

- amount of extra preparation required on the part of the incumbent;

- importance of this duty relative to the incumbent's other duties.

So much for the first part of the job description instrument, or "Instrument Part A".

(viewgraph 1)

Characteristics of the individual (assignee) and vocational satisfaction

Transfare the Task-No 100000

Check all the following questions. Mark the applicable answer No on the data acquisition sheet.

1. Highest educational level achieved

   Special elementary school ... 1

   Primary school without qualification examination or equivalent educational level 2

   Primary school with qualification examination or equivalent educational level 3

   Secondary school ordinary level or equivalent educational level 4

   Technical College with qualification examination 5

   Secondary school graduation - advanced level or equivalent educational level 6

   Higher professional school with qualification examination 7

   Technical College with qualification examination 8

   University education (diploma) 9

2. Status

   Conscript 10

   Long-term servicemen (21 months) 11

   Long-term servicemen (2 years) 12

   Long-term servicemen (3 years) 13

   Long-term serviceman (4 years) 14

   Long-term servicemen (5 to 7 years) 15

   Long-term servicemen (8 years) 16

   Long-term servicemen (9 to 11 years) 17

   Long-term servicemen (12 years) 18

   Long-term servicemen (13 to 14 years) 19

   Long-term servicemen (15 years) 20

   Regular soldier (20 years) .. 21

   Regular soldier 22

3. Career

   Administrative career 23

   Medical Service 24

   Military Band 25

   Military Geography 26

   Military specialist 27

4. How long have you been in your present position?

   Less than 3 months 28

   3 to 6 months 29

   7 to 12 months 30

   1 to 2 years 31

   more than 2 years 32

5. What is your present rank or comparative rank ?

Private (OR 1)	33
Private (OR 2)	34
Private 1st Class (OR 3)	35
Private 1st Class (OR 4)	36
Corporal/Petty Officer 3rd Class (OR 5)	37
Sergeant/Petty Officer 2nd Class (OR 6)	4o
Sergeant 1st Class/Petty Officer 1st Class (OR 7)	41
Master Sergeant/Chief Petty Officer (OR 8)	42
Senior Chief Petty Officer (OR 8)	43
Master Sergeant/Chief Petty Officer (OR 9)	44
2nd Lieutenant/Ensign	46
1st Lieutenant/Lieutenant	47
Captain/Lieutenant	48
Major/Lieutenant Commander	5o
Lieutenant Colonel/Commander	51
Colonel/Captain	54
Brigadier General/Rear Admiral (lower half)	55
Major General/Rear Admiral (upper half)	56
Lieutenant General/ Vice Admiral	57
General/Admiral	58

And now a few questions concerning your view regarding your position:

6. What is your opinion regarding your present activity ?

I find my present activity very interesting	59
predominantly interesting	6o
partially interesting	61
not interesting	62

7. Considering my training and qualification I am

satisfied	63
less satisfied	64
not satisfied	65
with my pay	

8. In my present activity my abilities are

hardly utilized	66
partially utilized	67
predominantly utilized	68
completely utilized	69

9. Considering the demands and performance called for in my present activity

I am

satisfied	7o
less satisfied	71
not satisfied	72
with my pay.	

1o. A different activity than my present activity would

surely give me more satis - faction	73
possibly give me more satis- faction	74
presumably not give me more satisfaction	75
definitely not give me more satisfaction	76

11. In comparison with similar activities in the Landesverhr my present activity is paid to low

yes	77
no	78

(viewgraph 3)

Controlling the camouflaging of sites and objects

Only if applicable → Transfere the Task-No 261701

Camouflaging of sites and objekcts

Only if applicable → Transfere the Task-No 261701

Check all the following questions. Mark the applicable answer No on the data acquisition sheet.

1. What do you camouflage ?

weapons and equipment	1
bunkers	2
land vehicles	3
air vehicles	4
water vehicles	5
fire positions	6
buildings, installations, facilities	7

2. Which camouflaging means do you use ?

terrain exploitation and selection of location (shadow, concealing cover, ground configuration)	8
natural camouflaging means (grass,foliage,bushes,branches)	9
protective coating against infrared and radar	1o
artificial means of camouflage (camouflage nets, tenting, artificial fog)	11
provisionary camouflaging means	12

3. What types of camouflage do you know ?

camouflage against

noise	13
visibility	14
light	15
infrared	16
radar	17

4. At what level do you accomplish this task ?

a)at the level of normal duties

several times daily	138
once daily	139
three times weekly	14o
one to two times weekly ...	141
one to three times monthly.	142
less than once monthly	143
never	137

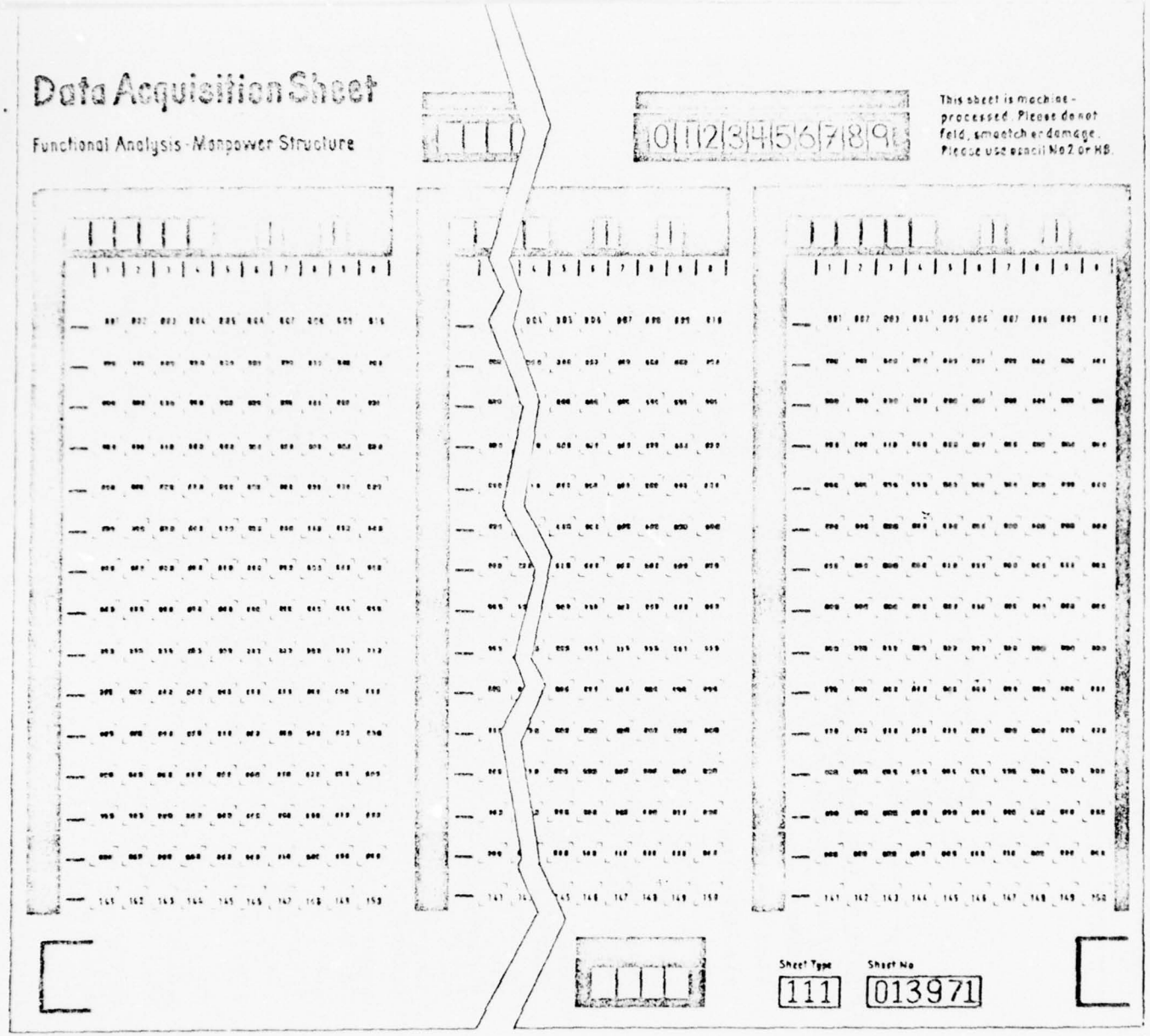
b)within courses or during exercises

more than 12 weeks annanually	132
up to 12 weeks annanually	133
1 to 4 weeks annnually ...	134
less than 1 week annanually	135
never	136

5. What significance has this task in comparison with other tasks of your position ?

select the applicable value on the scale between 1 and 7

1 very impotant main task ..	144
2	145
3 important task	146
4	147
5 less important task	148
6	149
7 comparatively insignificant secondary task	15o


Data Acquisition Sheet
Functional Analysis - Manpower Structure
0 1 2 3 4 5 6 7 8 9
This sheet is machine-processed. Please do not fold, smooth or damage. Please use pencil No.2 or HB.
Sheet Type
111
Sheet No
013971

(viewgraph 4)

Control of specialized and advanced training in the units	Only if applicable →	Transfere the corresponding task No from the ATB list (ATB means Training Activity Designation)
Specialized and advanced training in the units	Only if applicable →	Transfere the applicable task No from the ATB list. Check all the following questions. Mark the applicable answer No on the data acquisition sheet

1. How many persons from this activity do you provide specialized and advanced training simultaneously?
   - up to three 1
   - four to eight 2
   - nine to fifteen 3
   - sixteen to thirty 4
   - more than thirty 5

2. Is the training to be concluded by reaching a certain training level, attaining a permission, license or authorization?
   - no, no special qualification .. 6
   - attaining a permission/ license/ authorization 7
   - qualification certificate 8
   - reaching a training status:
   - status 8 (ORs) 9
   - status 7 (NCOs) 10
   - status 6 (Sergeants) 11
   - status 5 (Master Sergeants) .. 12
   - status 2 (Officers) 13

3. Do you provide on-the-job-training, i.e. during normal duty hours?
   - no 14
   - yes 15

4. Have you to perform practical exercises apart from your normal functions in order to provide personnel from this activity specialized and advanced training?
   - no 16
   - yes 17

5. Do you give theoretical instruction during the specialized and advanced training?
   - no 18
   - yes, daily 19
   - yes, two to three times weekly 20
   - yes, at least once weekly 21
   - yes, less than once weekly ... 22

6. In what form do you control the training status of this personnel?
   - no control necessary 23
   - oral tests (questions) 24
   - essays 25
   - written final examination 26
   - observation during performance of work 27

7. Does the specialized and advanced training of this personnel call for special preparations by you?
   - no, not necessary 28
   - provision of instructive and illustrative material 29
   - preparation of syllabus and sequence of syllabus 30

8. What significance has this task in comparison with other tasks of your position?
   Select the applicable value on the scale from 1 to 7
   - 1 very important main task .. 144
   - 2 145
   - 3 important task 146
   - 4 147
   - 5 less important task 148
   - 6 149
   - 7 comparatively unimportant secondary task 150

And now a few words concerning the second part, "Instrument Part B". In this part, not the performers of the duties themselves are questioned. Rather, after the data from survey part A have been analysed the duties in their various, ascertained modalities are submitted to experts to have them verify which knowledge and skills on the one hand and which abilities on the other hand are required for the performance of the respective duty in that certain modality. We are hence concerned with ascertaining the requirements that are associated with the performance of duties, that is requirements in terms of knowledge and skills that can be acquired on the one hand, and the relatively constant personality traits, hence the abilities which must be regarded as a prerequisite for the successful acquisition of a certain knowledge and of certain skills.

The testing of this instrument in the Bundeswehr began in June of this year. In this test the catalogues of questionnaires for the four specialty groups mentioned above, namely infantry duties, aircrew duties, navigation, and personnel administration, were used. Within the scope of survey part A in the four areas mentioned, some 3,200 servicemen at 60 duty stations, who had been selected in accordance with a certain selection plan, were questioned by 12 teams.

Extensive preparations were required. The briefing of project attendants of the Services as well as of the survey teams took a total of 4 days. In addition, it proved expedient to provide the survey teams with a 12-page guide to assist them in properly motivating the servicemen to be questioned.

In this pilot survey the servicemen questioned completed a total of 24,000 response sheets. On average, the incumbents questioned responded on 25 duties. The variance, however, was quite considerable. It ranged from a maximum of 98 to a minimum of 9 duties.

At this time the testing of "Instrument Part B" has not yet been fully completed. The findings from this test run will help determine the course to be followed by the Bundeswehr in its endeavor to reform its personnel structure.

Possibilities and Limitations of Job Analysis and Job Evaluation in Armed Forces

Lieutenant Colonel Seuberlich
German Army

Let me begin with a preliminary remark. At one of the many meetings which of late have been held to discuss analytically and critically questions of performance, the Minister of Economics of the Federal Republic of Germany said, in May 1973, among other things that:

The performance motive of some was sometimes considered as a pretext for the profit and power motive of others. It also served as an expression of greed and the striving for power, as few gave orders and many obeyed. Also, individual work performance in a highly industrialized economy was not, or only indirectly, measurable. Performance was not measured by the market value of labour but by the market value of the product. This inability of performance to be measured was apparent, among other things, from the fact that wages and salaries were mostly fixed by collective bargaining, that is independent of individual performance. Replacing the performance principle by some other principle, however, would not remove the negative aspects for which, incidentally, the performance principle was not the cause. Independent of the collectively negotiated wage structures, which had to be retained in the economy, there was the possibility for the individual to be rated on his performance. However, the Minister conceded, an objective evaluation mechanism was not yet known. In this connection it was also necessary that psychologically-motivated performance readiness be accompanied by performance ability, which in turn presupposed equality of opportunity for all, which again could be provided only by an efficient educational system.

From all this it is already apparent to what extent the "mystery that is man" - mysterious with respect to the interactions between requirements, performance and evaluation - constitutes the determining factor, and this factor will be the principal point of reference throughout my presentation. When I say "the mystery that is man" I use that expression to allude to the difficulty of predicting human behaviour and point to one of the reasons why, for instance, entrepreneurs concentrate their planning and goals predominantly on the management of inanimate things. As soon as man, however, becomes the object of planning, psychological, work-physiological, and sociological aspects become dominant. They confront planners with a host of bewildering determinants for every planning model. The problem here lies in the ascertainment, quantification and also qualitative determination of all the essential factors which influence the behaviour patterns of man and thus make possible prognostication and planning.

With regard to the questions that pertain to armed forces we have to keep one thing foremost in mind: the behavior of man cannot be predicted with absolute certainty; it is never static but a correlating product - subject to constant change - of the personality structure of the individual and the relevant environmental situation in which he finds himself. Of course, nobody today would deny that there are test instruments with a sufficient degree of reliability, and meaningful tables of norms which have proven themselves to be useful instruments of prognostication. By using such instruments the Hamburg police school, for instance, was able to reduce the number of their police school dropouts from an initial 10 percent to 3 percent. And in the endeavors to develop the basis for a new personnel structure in the German Armed Forces the problem of performance prediction will soon also gain central importance. If the means of psychodiagnosis are to be employed merely for the purpose of detecting extra educational capacity in efforts to control the development of talented individuals and to systematize the advancement of co-workers, this does not, in principle, pose any serious difficulties since, it is true, a time span of 45 minutes is sufficient to determine with adequate accuracy the intelligence quotient of 35 individuals. But here the question already arises as to what is actually expressed in the intelligence quotient? What kind of intelligence is measured?

By a definition dating from 1923 intelligence is simply what is measured in the intelligence test. This definition, as is widely agreed, has long been overtaken by developments. More acceptable to us today is a definition from 1963 according to which "intelligence is the ability of an individual to think in concrete or abstract verbal, numerical and space-time terms; it enables him to deal successfully with many complex - and with the aid of particular groups of abilities - also quite specific, situations and tasks."

This definition already takes into account the findings of modern intelligence research which holds that intelligence cannot be expressed in a rigid number. And what is more, intelligence here is expressly related to the successful dealing with complex and also very specific situations and tasks. And this brings us to the heart of our subject.

As early as 1969 the Manpower Structure Commission appointed by the Federal Minister of Defense, determined as a result of its first investigations:

The concept of the standard officer and standard non-commissioned officer with a broad utilization spectrum and broad interchangeability can no longer be kept up on a modern personnel structure. This obsolete concept must be replaced with a modern, vocation- and function-related concept. In particular it must, through service-peculiar differentiations, be in consonance with the multiple division of labor in modern armed forces.

The new vocation- and function-related concept of the modern soldier will have repercussions in many areas. It will, among other things, have to stress a duty-related typology of the soldier. Quite revealing in this respect is the attitude of the Air Force who hold that in the Air Force everybody today is a specialist. In this connection it is interesting to note the following approximate distribution: only 5 percent of all jobs are directly combat-oriented while 95 percent may be subsumed under the headings of technics, economics, or administration. These latter jobs can also largely be compared with corresponding civilian jobs. But regardless thereof, standard criteria must be developed for the representation and categorization of the great variety of occupations in the armed forces. Special attention must in this context be given to the adequate recognition of civilian vocational background knowledge and experience, especially for appointment, advancement and career progression purposes.

This finding is based above all on comprehensive polls of young men of draft age. It appeared that their attitude toward the Bundeswehr was influenced in a negative way as soon as they included their own situation in their considerations. This applied to those who took a generally positive view of the Bundeswehr as well as to the group whose overall attitude to the Bundeswehr was negative to begin with. It is worth noting that this trend also exists among potential NCOs although this group shows, on the whole, a much more positive affinity with the Bundeswehr. The greatest shift in attitude was noted among high school graduates. Attitude research confirmed that it is above all practical considerations which determine their actions, for instance their decision to enlist beyond their statutory term of service. From this it is evident what consequences a negative shift in attitude as a result of his practical considerations can have for an individual's actual behaviour. These practical considerations include in particular the career prospects that potential draftees think the Bundeswehr can offer. Only one fifth of those polled rated their career chances in the Bundeswehr as clearly positive. The majority of the respondents who by their education were eligible for an officer career do not believe, for instance, that to become an officer requires special qualifications. Although of the two target groups (potential NCOs and potential officers) nearly one third of those polled (32 percent and 27 percent respectively) do in principle consider volunteer service in the Bundeswehr to be a possible career alternative, this is not saying anything about the actual recruitment potential; what is decisive, rather, are the concrete reasons for which potential NCOs and officers would be willing to volunteer. Here it is found that rational considerations determine their readiness to volunteer. Financial security (45%), good vocational training (35%), good chances of career progression (18%) range foremost among the positive motivations whereas "liking the military profession" motivates only 7 percent.

One of the central problems of personnel management is hence utilization consistency. This is the degree of congruity of knowledge and abilities on the one hand and the job with its concrete duties on the other. A high degree of congruity not only increases work efficiency but also enhances the individual's subjective contentment in his job. An analysis of utilization consistency in the Bundeswehr showed a notable discrepancy with utilization consistency in the civilian sector. Only 12 percent of the conscripts and only 26 percent of the temporary volunteers feel that in their present jobs in the Bundeswehr they are employed according to their qualifications and abilities. Only in the group of regular soldiers is the degree of utilization consistency much higher: no less than 56 percent consider being employed in keeping with their knowledge and abilities. But even here the discrepancy with the corresponding figure in the civilian sector is still 14 percent, the consistency value in the civilian sector having been established at 70 percent.

In evaluating this situation we must of course keep in mind the particularity of the mission of the Bundeswehr and as a result thereof consider that military service has certain characteristics which although they may also be encountered in other areas are - because of their combination and weight - nevertheless significant for the armed forces. Soldiers must not only accept certain abridgements of their individual freedoms; they must also daily prepare to use force and learn to operate into situations of danger. These particularities which are function-determined and irrevocable are also to be clearly shown by job analysis as being indispensable for the mission, in order that they may be duly taken into account in the assigning of pay and service grade.

This is a difficult task to accomplish, given the present pay and grade rules in effect in the Bundeswehr, rules, which are oriented chiefly on traditional criteria and which have only partially been adjusted to changed internal and external conditions.

Among the major structural elements of the Bundeswehr which make solving its personnel problems difficult is not just the already-mentioned, out-dated concept of the standard officer and standard noncommissioned officer, the very principle of which is at variance with the growing differentiation of job requirements. Directly associated therewith is the system of a summary rather than a requirements-related, i.e., analytically determinable, job evaluation. And finally there is the structure of the position pyramid which also does not allow for the necessary degree of differentiation but which is rather also in the nature of a "standard position pyramid".

Thus the main elements of the present personnel structure are characterized by criteria which operate to make the filling of existing positions with adequately qualified personnel and the assurance of an adequate performance readiness on the part of recruited personnel a latent problem.

Let us now cast a brief glance at the complex of evaluation and pay. Here the lack of precise military job profiles entails a variety of disadvantages since the present evaluation and pay system is based on rank. Accordingly it is, by its nature, linear in structure and relatively little adaptive. Thus, the major tools of a modern personnel structure policy, i.e., the assigning of pay and service grades, are severely restricted in their potential. But nothing than the structure of a specific job profile, i.e., the presence of specific job characteristics in a certain constellation, can be the criterion for the conjunction of positions and grades.

In some specialty areas the discrepancy between job and service grade was ascertained. From the investigation it was obvious that a redistribution of many enlisted men presently in NCO grades to other grades was necessary. Thus, for 23 percent of the NCOs in the specialty area of administrative duties the syndrome of duties comes closer to that of lower ranks than to the job profile of higher service grades. This was only an example.

In connection with these analyses another finding deserves to be mentioned. The subjective importance of improper grading with respect to its effect on job satisfaction is different from what it had been suspected to be. Neither are soldiers in higher-rated positions especially content with their pay nor are soldiers in lower-rated positions especially dissatisfied. This result confirms that satisfaction with pay is determined rather by factors of time pressure and work load and, overall, the degree of strain involved in a job. In the problematical group of NCOs, for instance, much more than in any other group, utilization in higher-qualified but not correspondingly higher-paying positions is translated into discontentment with their chances of career progression.

In evaluating this problem area we must also consider the factor of job satisfaction in conjunction with the utilization consistency of personnel. It has been shown that the factor "fighter activity" exerts the strongest influence on utilization consistency, because with the growing content of fighter duties in a job less and less civilian vocational skills and abilities can be applied. Here the peculiarity of the mission of the Bundeswehr is clearly apparent. There simply is no civilian equivalent for fighter duties in the stricter military sense of the word. This means that by and large only a small percentage of personnel in this area find that degree of job satisfaction which is required for a positive concept of self. It is above all the type of individual who quite unreflectedly thinks that "being a soldier is fun". The negative influence exerted by fighter duties on utilization consistency can, however, be largely offset by the strain imposed by extreme work conditions as well as by a greater measure of communicative activity, especially if these two factors can be combined. In addition, activities of an organizing and planning nature are apt to reinforce utilization consistency. All in all, this is an area which requires further investigation.

Also to be mentioned in this context is the attempt by the Bundeswehr to regard discipline and personal freedom as not being mutually exclusive. It was undertaken to redefine, under the term "Innere Fuhrung" (inner conduct and leadership), the concept of command and leadership under the changed conditions obtaining when the new German armed forces were organized beginning in 1955. Here especially three factors come into play which are sufficiently relevant also in the context of evaluation to be briefly mentioned here:

1. A certain apprehension that German militarism might possibly be revived was dispelled by the lawmaker and the Bundeswehr leadership when it was decided that man as the center of our political and social system should, also as a citizen in uniform, remain at the center of our defense efforts.

2. Along with the rest of technology, arms technology had advanced so dynamically that also for this reason new modern forms of training, leadership and tactics were needed. Thus, the situation was and is that no superior can acquire the whole range of diversified special knowledge. Rather, he has to rely on the conscientiousness and competence of qualified subordinates.

3. The threat to the free world was not restricted to the threat of force. Therefore, also the moral struggle had and has to be considered, in which the soldier needs knowledge, judgment and inner strength to sustain him.

From this point of view, every military superior is also an educator in a special area. This is a constant, central task since even in peacetime spiritual and moral forces often determine the value of a soldier more decisively than do specialized skills. Understanding for the duties and the conscientiousness with which they are performed determine the state of training of the troops and the operational readiness of the weapon systems. Thus, the deterrent effect of the Bundeswehr depends at all times on the sum of the activities which the soldier performs in peacetime. In the Operations Manual of the German Army it says: "War is the severest test of man's moral and physical strength. In war, moral qualities are therefore often more important than intellectual abilities . . . The vastness and emptiness of the battlefield, the tremendous effect of firepower, the rapidity of all maneuvers and the absence of fixed fronts more than ever call for soldiers who show individual initiative and responsibility and who act in furtherance of the overall objective." This quotation clearly describes the aims for the soldier's education in the Bundeswehr. The importance of these aims in all armed forces throughout the world is uncontested. But their attainment is possible only to a limited extent and hardly less difficult than a systematic analysis of the respective requirements to be met by the individual and their corresponding evaluation. The psychological factors which lie in conflict here are too heterogeneous. Consider for instance the situation when panic broke out 1914 in the battle near Gawaiten-Gumbinnen as described by Hesse in his "General Psychologos".

The Army Operations Manual calls this phenomenon a sudden breakdown of moral resistances. It can be brought on by overwhelming combat experiences as well as by seemingly insignificant or unrelated events, but also by psychological measures of the enemy. Stouffer, the well-known American military sociologist comes to the conclusions that personal motives and relationships do not alone determine the combat effectiveness of fighting units. Officers and enlisted men must be motivated to that end. The best criterion on which to base a prediction of behavior in combat, according to Stouffer, is the individual's attitude toward his institutionalized role which says that he is a soldier, not a civilian. The social role confronts him with the alternative of either fight or bear the institutionally envisioned consequences.

But we know that in spite of intensive efforts no system has as yet been devised that allows us to predict which soldiers would actually fight and which would just trudge along. Behind all this is the question of defense motivation which is of vital importance; for the readiness to commit oneself up to the point of self-sacrifice is not a matter of course. The "what for" must be strong enough to overcome the existential anxiety in man. The "what for" must also provide the impulse for the individual's physical commitment and loyal performance of duty in the daily routine of peacetime military service. For the individual soldier the meaning or senselessness of his actions can be derived only from the "what for". Only a strong "what for" motive makes for conviction and leads to an insight into and acceptance of the special conditions of military life that are so different in many ways from those of a civilian existence.

The requirement for initiative and responsibility on the part of the individual, down to the last man who is a specialist in a certain function, follows from what has just been said.

With the above aspects in mind the author developed a procedure by which he tested 1,200 conscripts immediately after they were inducted for military service in order to

- determine how the aptitudes, attitudes and abilities of raw conscripts correspond to the special requirements of service in the Bundeswehr;

- ascertain the characteristics by which groups of various educational attainment differ from each other;

- establish criteria for the formation of realistic training classes;

- ascertain relevant characteristics for grouping individuals in different training classes, considering individual aptitude patterns.

At first, the overall objective was to find a military intelligence quotient. This goal, however, had soon to be abandoned as the number of variables was too great and as too many different evaluation factors had

to be applied so that in operationalization it proved impossible to combine and express the necessary differentiated statements in one numeric quantity. On the other hand, it turned out to be quite feasible and also much more expedient to express the findings and values in a military intelligence profile. The test procedure was adapted to the specific situation of the young people prior to their induction for military service.

The test was conducted during the so-called "enrollment processing phase" of the recruits, that is to say during a period when the influence exerted by instructors was still minimal. Hence, the test reflects almost exclusively the raw recruit's original equipment in terms of aptitudes, skills, knowledge and views.

The test was divided into two main parts. The academic part comprised 40 main questions, each containing up to 20 subquestions. The questions had to be answered under time pressure and were arranged in such erratic sequence that a pattern could not be recognized by the testees. The practical part contained twelve main tasks, each with up to six sub-tasks, some of which had to be solved at night in unfamiliar terrain. In the evaluation the results of the academic part and of the practical part were interrelated in a multiplicity of ways and then consolidated into 18 overall grades for each individual soldier. In all, the five following main areas with their subdivisions were evaluated (viewgraph 1):

Viewgraph 1

1. Physical abilities
   - sports performance
   - taxability and endurance
   - reaction
2. Mental abilities
   - technical understanding
   - orientation, and grasp of terrain
   - concentration and receptivity
3. Personality traits
   - purposiveness
   - readiness to commit self and to assume responsibilities
   - decision
   - leadership
4. Relationships with the environment
   - attitude towards fellow men and companionship
   - prior relationship with superiors (parents, teachers, principals, etc.)
   - concept of military superiors
5. Civic knowledge and attitudes
   - civic knowledge
   - perception of the Communist threat
   - attitude to the notion of fatherland
   - attitude to the notion of a free democracy
   - attitude to military service.

The consolidated test results show a military intelligence profile, developed and measured on conscript recruits inducted in 1966 and broken down by educational groups, and presented in the 18 areas mentioned. I will now show you two segments (viewgraph 2):

The grade scale contains values 1 - 3, each with nine intermediate values. The grades denote:

1.0 = fully suitable
1.5 = suitable
2.0 = conditionally suitable
2.5 = not yet suitable
3.0 = unsuitable

The horizontal line at 2.0 indicates the borderline between suitability and conditional suitability; it may be called the military minimum requirement which the young soldier must have exceeded by the end of basic training in order to meet the subsequent requirements.

The educational groups are identified by letters as follows:

VB = elementary school leavers (9 years of schooling with subsequent vocational training) (303)

A = high-school graduates (13 years of schooling) (100)

M = high-school and commercial school leavers (10 years of schooling) (100)

V = elementary and special school leavers without vocational training (31)

D = average of persons surveyed (534)

Mental abilities (viewgraph 2)

With a score of 2.0 the technical understanding of the high-school graduates is already suitable, that of the unskilled workers (elementary school leavers without vocational training) on the other hand, at 2.4, still very much in need of development. In technical understanding the conscripts are, on average, brought up relatively quickly to the requirements of training.

Also in the area of "grasp of terrain" (orientation in and perception of terrain, e.g., estimating distances, utilizing terrain features, selecting a position) the group of high-school graduates, at 2.0, distinguish themselves by a suitable performance and also during the subsequent training showed themselves generally more capable of development than the soldiers of the other groups. But even for these the distance to the dotted line can be closed within a few weeks from 2.2 to a satisfactory 2.0.

The poor test performance in concentration and receptivity is symptomatic of the generally noticeable fluctuating intelligence even of high-school graduates, caused by the flood of stimuli of the modern urban world which with television, cars and neon signs has invaded even the remotest villages. While it is possible to eventually close this gap it requires a lot of patience, good didactics and toughness on the part of the instructor. With the average performance being 2.5, this is not so much a question of intelligence than of education and the environment. Even the high-school graduates, having scored 2.1, are initially on average only conditionally up to difficult concentration tasks. A rating of 1 was achieved by only 4 percent.

Now as for <u>personality traits</u> (viewgraph 3):

To determine purposiveness in a short test is possible only to a very limited extent since only disposition and tendencies but not the decisive factor, namely perseverance of will over an extended period of time, could be tested. Seen in this light the test result, as shown in the viewgraph, is only of limited value. On the other hand, the high average score of 1.5 reflects our "open, permeable" performance-oriented society. In other words: in our society it is worthwhile for everyone to set goals for himself. Whether he attains these goals depends primarily on his own capabilities, his efficiency, knowledge, perseverance, etc. - in short, on his own performance. The degree to which he uses his potential is a question of intelligence, energy and endurance. It deserves to be noted that this fact is apparently also well known among unskilled workers (elementary school-leavers without vocational training) although they, probably as a result of their past experiences, show less "courage" to set themselves goals than the other groups.

In summary, the test result permits this conclusion to be drawn which is of particular importance for the military leader: the young soldier is receptive, he is responsive; it depends essentially on the skill of the military superior to keep this responsiveness alive, to foster it and to transform it into willing independent cooperation.

The readiness of an individual to commit himself and to assume responsibility, too, can be determined in a test only to a limited extent although in this area substantially more secured information could be gained through practical tasks. But the degree of reliability of the information must still be rated as highly "conditional" as in the final analysis no one can say with absolute certainty of himself how he would react sometime in the future in some situation or other under variegated specific conditions. His behaviour depends ultimately on a great number of factors and circumstances, not the least among them being the individual's "condition of the day" if we want to call it that, which may be determined to a larger or lesser degree by hunger, fatigue, physical injuries, etc. Nevertheless, the average score of slightly over 2.0 indicates a suitable basis. Among the young conscripts the readiness to assume responsibilities and to commit self is widely distributed. It is fully sufficient as a basis on which to build and secure training and education.

With an average rating of nearly 2.0 the decision-making power of the conscripts may, from the very first day, be called sufficient. The graph shows three interesting peculiarities:

- The group of the high-school graduates does not rank first. It obtained the same average mark of 2.0 as the high-school and commercial school leavers with 10 years of schooling and the unskilled workers. The reason may be seen in the fact that because of their greater mental agility high-school graduates are aware of many more possibilities than their lesser educated peers. For the high-school graduate the result is a certain hesitancy - he weighs and considers, often much longer than the others, before he finally makes his decision.

- The group of the elementary school leavers with subsequent vocational training is the only one which, with a mark of 2.1, stays below the dotted line. They are the only group among the testees who have both a completed vocational training and, on average, comparatively longer experience in jobs of somewhat increased responsibility; in other words they are a little more aware than the others of the practical burden of responsibility, and this is perhaps why, on average, they hesitated somewhat more than the others in making their decision.

- The unskilled workers seem to have approached the decision tasks pretty "nonchalantly", and in doing so scores astonishingly well.

In summary it may be said: mind, education, and decision are interrelated in a manner which is hard to objectivate. In the lower range of military requirements for decision-making the level of education and knowledge plays only a minor part. The natural aptitude of the average conscript in this area fully meets the requirements.

What has been said of decision applies analogously to leadership. In all educational groups there are conscripts in sufficient number who possess enough original aptitude for minor leadership tasks. A certain superiority of the group of high-school graduates and high-school and commercial school leavers with 10 years of schooling, however, is already apparent. Practical experience during training bears out this finding. The superiority of these two groups increases as leadership demands get higher. This does not mean, however, that individual other soldiers, especially among the group of elementary school leavers with subsequent vocational training, could not also achieve the same results. The top performers in the VB group often even surpass the average high-school graduate.

Let me add that leadership qualities were tested, among other things, in a simulated crisis situation, for groups of five testees each in order to determine - to the extent that this is possible at all - social spontaneity in group-dynamic interaction.

But let us now consider the possibilities of evaluation on the basis of three general criteria:

A. Jobs of the executing type, that is jobs involving predominantly executing responsibility.

B. Jobs of a predominantly - "tactical" - nature, as I will call it, characterized as a rule by a mixture of executing and directing responsibility.

C. Jobs of a predominantly "strategic" nature, characterized by an outstanding degree of directing responsibility.

In jobs of group A incumbents are told what has to be done and how to do it. An incumbent of this type of job must stick as closely as possible to regulations in order to satisfy the requirements. Although he may suggest changes he is not authorized to direct any. Incidentally, job descriptions for positions of this type can be drawn up by what might be called job description technicians in accordance with established guidance or even by the incumbents themselves. Most job descriptions, as we know, are of this type.

Not so in the case of group B. In this category belong all those positions or jobs of the lower or middle management and of specialists whose main task is to plan, direct and supervise the activities of the executing incumbents. Generally, therefore, only the results to be achieved are prescribed for the B-group. Hence, incumbents are expected to work out for themselves whatever measures are needed to achieve these results. The content of such a position is hence determined to a considerable extent by the incumbent himself. Within certain limits he can freely apply his specific knowledge and faculties. Therefore, methods of job description and job evaluation developed with a view to the A-group are of necessity no longer fully applicable to jobs in the B-group. In the jobs of this group the multiplicity, heterogeneity, complexity, and changing priorities of the duties interact in such a way that the method of listing duties and tasks is no longer sufficient.

The jobs of a predominantly "strategic" nature, finally, are top positions. Characteristically, these are the positions where it is determined what is to happen to an organization in the long run, what aims it is to pursue, where it is headed. A strategic position is therefore shaped largely by the incumbent. Here the complete unfolding of the incumbent and his identification with the job must be expected and demanded. Therefore, the mere description of a strategic position alone requires an extra-ordinary degree of empathy with management functions and understanding for problems involving a high degree of directing responsibility. Description, analysis and evaluation of such jobs is for the majority of presently applied methods a test which only very few are likely to pass.

In spite of the difficulties indicated even large organization of the size and complexity of modern armed forces, which in the area of personnel organization have embarked on job description, analysis, and evaluation, are likely to try to include the entire organization in a uniform description and evaluation system. If only because of the high personnel fluctuation on the one hand and of rotation on the other any vertical or horizontal limitation of whatever kind in the system must be ruled for the armed forces. Otherwise a marked decline in urgently required performance motivation would have to be expected since the requirement for structural visibility and transparency could no longer be adequately met.

For all the reasons mentioned the Geneva system of work evaluation, at least as far as the four basic requirements are concerned, should be applied in principle also to job evaluation in armed forces. In pursuance of this aim the Bundeswehr has commissioned within the framework of its basic personnel structure studies a "job and employment standard" (TVS) to be developed as a specific structural instrument of organization. The job standards are designed to cover the typical duties of the positions in their typical modalities as well as the associated requirements profile. They thus constitute the descriptive part of the TVS. They can be grouped and dealt with under such aspects as personnel structure, training, personnel management and personnel regeneration. This is the first step. In a next step the job standards are then compared with each other for similarity and with respect to the level of required qualifications in order to develop criteria for the establishment of a system of specialty groups. Here the statements are each referred to a job standard and presented in an employment standard. The combination of the two statement complexes results in the job and employment standard, the TVS.

The importance of job and employment standards extends into all areas of personnel organization, planning, management, and training, because in all areas problem solutions must be based on data and assumptions concerning the contents of jobs.

The conceived TVS consists of 12 parts. Parts 1 and 2, with their job designation (ATB) and other identifying references are to enable general classification, that is the systematic arrangement of individual positions. Parts 3 and 4 serve the purpose of ordering duties according to executing and directing responsibility and by centrality. Parts 5 and 6 list the principal abilities and skills required and the necessary school and vocational education. Part 7 describes the required career background of incumbents, differentiating between civilian and military prior employment. Furthermore, reference is made to training courses which are considered either to be beneficial or required for successful performance of the job. Parts 8 and 9 show, especially for first assignments, the rquirements in terms of aptitude and suitability and for subsequent assignments in terms of capability and experience. Part 10 deals with

the vertical and horizontal military and civilian assignment potential (the so-called subsequent career). Parts 11 and 12, finally, contain data which are derived from duties and requirements, such as grade of pay, minimum time in a job, etc.

It is envisaged that the TVS contain a maximum of twelve requirements characteristics whose occurrence in the qualification profile of an individual contribute to the successful performance of the duties of the standard. These characteristics must be rank-ordered. The rank-order is derived from the various criteria of the duty and from the requirements analysis. As for the subsequent career, the determination of the possibilities of horizontal and vertical employment of personnel is based on the same methodical premises as the determination of the preceding career.

I now come to the conclusion of my presentation:

The objective of function analysis within the scope of the groundwork for a new personnel structure of the Bundeswehr calls for standards whose statements are so differentiated that all associated positions can be uniformly dealt with under the aspects of personnel organization, personnel management, personnel planning and training. It follows from this requirement that by reason of a factually correct rank-ordering of the individual position also the evaluation of that position and thus its pay grade is fixed. The grade of pay will ultimately be determined, however, not only by the characteristics of the job but also by the experience and performance of the individual incumbent. Rank, by contrast, cannot be geared to the person or his performance but must exclusively be determined by the characteristics of the job, namely by the requirement for visible expression of directing responsibility.

We know that a long and difficult road is ahead of us until this objective is achieved.

Occupational Analysis in the Royal Australian Air Force

by

Captain Wayne S. Sellman and Squadron Leader John W.K. Fugill
Headquarters Support Command
Royal Australian Air Force
Victoria Barracks
Melbourne, Victoria

Job analysis and its related procedures have become fairly common practices in many organizations as a means of providing personnel planners and managers with information about the nature and requirements of on-going work. Because of sharply rising costs and the seemingly inevitable concomitant manpower reductions, such occupational information has recently assumed an even greater importance within military services around the world. More efficient personnel selection and recruitment, classification and placement, training, personnel appraisal, job evaluation, and job design are absolute necessities as the size, composition, and function of the various armed forces come under closer and closer scrutiny by their governments.

Although the Royal Australian Air Force (RAAF) has long recognized the need to know more precisely what its airmen do on their jobs, it has only been in the past five years that a systematic endeavour for collecting, analysing, and reporting job information has been undertaken. Based primarily on techniques reviewed overseas during an investigation of Royal Air Force, Canadian Defense Forces, and United States Air Force occupational research programs, the RAAF, in 1967, initiated a modest job analysis and training evaluation effort. During the next two years, two job analyses were accomplished. Unfortunately, these studies were conducted by a job analyst as an additional duty rather than as a primary job, and in addition they suffered from being different in design and approach, with the survey instruments, response forms, and rating scales varying on each occasion. This lack of standardization proved especially costly in both job analyst and computer programmer time as unique EDP programs were required to analyse each return. However, the situation was understandable as these surveys represented the first attempts to introduce job analysis into the RAAF and was the natural result of trying to find the 'best way'.

In the meantime, the RAAF was seeking closer co-operation with the USAF through the formulation of an exchange program in the job analysis area. This exchange became a reality in June 1970 with the assignment of a RAAF squadron leader (major) education officer to the Air Force Human Resources Laboratory in San Antonio to study occupational research techniques and operational job analysis procedures. At the same time, the USAF exchange officer, a captain psychologist, was posted to Headquarters Support Command in Melbourne to develop and supervise the implementation of appropriate job analysis and training evaluation methodologies within the RAAF.

/Thus

Thus, the RAAF system of job analysis described in this paper is a modified version of the USAF job inventory approach pioneered over the last decade by the United States Air Force Personnel Research Laboratory, now a division of the Air Force Human Resources Laboratory.

RAAF Occupational Survey Procedures

Although the job inventory approach to occupational analysis (Morsh, 1969; Morsh, 1964; Morsh, Madden, and Christal, 1961) is well known to most of you, we would like to briefly summarize the RAAF version. First there is the construction of the job inventory. As you know, the job inventory is composed of two basic sections: background information questions, and a list of the significant tasks performed by airmen at all levels of rank and job proficiency within a mustering (USAF specialty). An initial listing of tasks is constructed from available published sources such as manuals, regulations, training documents, etc. This list is then subjected to an interview review by senior NCO subject matter specialists, and to a mail review by additional senior NCO specialists working in the field. The preliminary listing of tasks is then revised in light of their comments and recommendations for changes, additions, or deletions, and the final version of the job inventory is prepared.

The inventories are then mailed to unit trade test officers (analogous to USAF test control officers) who administer them under standardized conditions to selected groups of job incumbents in base testing rooms. We should note here that great care is taken to select representative samples. In the larger musterings (N 800) samples between 50 and 75 percent are used while 100 percent sampling is the goal for the smaller musterings.

In completing the inventories, job incumbents provide identification, background, and other job related information. Then using specially designed machine-readable answer sheets, they respond to the task statements by indicating each task performed and the relative amount of time spent on it. Finally, incumbents add tasks not listed in the inventory. Because the task response answer sheets used are to be optically scanned, incumbents write-in their additional tasks in the background information section, not on the task answer sheets themselves.

Completed inventories are returned to us at Headquarters Support Command where they are checked for errors. Background information is forwarded to the RAAF EDP Centre for keypunching and transfer to magnetic tape. The task responses go to a local university where they are converted to tape by an optical scanner. (By the way, the optical scan error rate for our last three surveys was less than ½ percent.) The task response tape is then sent to the RAAF EDP Centre where, by means of subject respondent codes, the two tapes are merged into one which contains the complete file for each airman in sequence. The data are then re-edited and prepared for computer analysis.

One of the major advantages of the job inventory approach to occupational analysis is that the results are comprehensive and easily quantified (Christal, 1971). To fully exploit this advantage, we are working with the RAAF EDP Centre in the development of a catalogue of standard computerized occupational analysis programs. To date, we have available two programs for application to our job surveys - group job descriptions, and group difference descriptions. The group job description program enables the computer to generate job descriptions, made up of tasks performed and the

/average

average percentage of time spent on each of them for any group of individuals where the groups can be defined in terms of background variables. For example, a job description can be computed for all airmen of a given rank, or for those assigned to a specific work area, or for those who have been on the job for less than six months. Indeed, a group job description can be generated in terms of any single or combination of background variables. Thus, a group job description could be provided of all corporal, electrical fitters, between the ages of 20-25, with over three years experience in the mustering, and over one year in their present position. The background variables and possible combinations on which groups may be categorized are limited only by the job analyst's imagination.

Our second program provides comparisons between groups. These group difference descriptions effectively point out the difference between two groups in terms of the emphasis they place upon various tasks. With such descriptions, it is possible to determine how the time spent on tasks varies with rank, time in the job, time in service, by unit, by area of assignment, or any other background variable completed by the incumbents. For example, a comparison by rank would depict how the job changes as one progresses in rank; or which tasks are performed by junior members and which by more senior ones. A comparison of major work locations would indicate which tasks are performed by airmen assigned to flying squadrons as opposed to those which are performed by aircraft depot members. Greater detail concerning the specifics of the programs can be found in the writings of Archer (1966); Morsh and Christal (1966); and Christal (1971).

Those of you familiar with the job inventory approach will immediately recognize that the above described programs are versions of two CODAP (Comprehensive Occupational Data Analysis Programs) programs designed by Dr Raymond E. Christal of the Air Force Human Resources Laboratory. While the RAAF does not possess any of the other CODAP programs at present, it is studying the feasibility and cost-effectiveness of using the programs in a relatively small force.

Some Examples of RAAF Occupational Data

In the USAF, the basic unit of the airmen occupational classification structure is the specialty. Specialties group related work positions on the basis of knowledge, education, training, experience, and other abilities required to perform them. Specialties are grouped into larger clusters called career fields. This clustering is based on the similarity of the specialties and their transferability of skills and knowledges to provide optimum utilization and career development. The USAF presently uses some 46 career fields encompassing approximately 235 specialties. Skill distinction and progression within a career field is indicated by a vertical arrangement of specialties. Thus an airman moves 'up the career ladder' from apprentice to journeyman to supervisor to superintendent. Each skill level within a career ladder is designated as a separate specialty.

Within the RAAF, airmen are assigned into 25 primary work areas known as trade groups (similar to USAF career fields) which embrace a total of 120 different musterings (USAF specialties). Each trade group is organized along as functional lines as possible. For example, the catering trade group contains such musterings as cook, cook's assistant, steward, caterer, and butcher. In some instances, a mustering is used as a 'feeder' for a more specialized mustering within the same trade group thus establishing a career plan. Thus, a cook's assistant, by further training and testing,

/could

could progress to cook. This type of career progress is somewhat analogous to the USAF's career ladder system. It differs in that while airmen in the USAF are expected to progress in terms of specialization, RAAF airmen may remain in a mustering with little upward mobility. The scope of RAAF work activities, therefore, is accomplished by 120 airmen specialties while the USAF uses over 235.

Because of the smaller number of musterings, you might expect that RAAF airmen would be involved in a much wider variety of job activities than are their USAF counterparts working in similar specialties. While we have no USAF data on comparable specialties available at the present time, results from our first two occupational surveys (Sellman & Crow, 1972; Sellman & Crow, 1973) suggest some support for this notion. Perhaps some examples of RAAF data at this point might be illustrative. Tables 1 and 2 present for the total survey samples and the major operational work areas the extent to which work time is spread across tasks. This is done in terms of the number of tasks accounting for 10 percent, 25 percent, 50 percent, 75 percent, 90 percent, and 100 percent work time as tabulated from group job descriptions. It can be seen that for major work areas, group work time is spread across a larger number of tasks for the operational locations such as aircraft depots, flying squadrons, and maintenance squadrons than for the non operational areas such as the training schools, command headquarters, and the Department of Air. The number of tasks accounting for work time can also be considered a measure of job specialization - the more tasks performed the lower the specialization, and vice versa. For example, training schools and DEPAIR personnel are highly specialized in their particular jobs because they spend almost all their time performing a few activities; aircraft depot airmen, on the other hand, are less specialized in that they spend their working hours performing a large number of tasks.

TABLE 1

NUMBER OF TASKS ACCOUNTING FOR WORK TIME OF ELECTRICAL ENGINEERING MAJOR WORK AREAS AND TOTAL SURVEY SAMPLE

		Number of Tasks Accounting For					
Group	N	10% of Time	25% of Time	50% of Time	75% of Time	90% of Time	100% of Time
TRG SCHOOL	20	1	5	19	58	126	273
DEPAIR/CMD	19	1	4	13	33	61	133
MNT SQN	199	19	58	151	310	507	1146
FLY SQN	146	13	40	112	255	419	899
ACFT DEPOT	95	12	39	119	269	446	1120
TOTAL SAMPLE	496	17	54	147	308	491	1148*

*1148 Tasks Listed in the Electrical Engineering Job Inventory

/TABLE 2

TABLE 2

NUMBER OF TASKS ACCOUNTING FOR WORK TIME OF INSTRUMENT ENGINEERING MAJOR WORK AREAS, AND TOTAL SURVEY SAMPLE

		Number of Tasks Accounting For					
Group	N	10% of Time	25% of Time	50% of Time	75% of Time	90% of Time	100% of Time
CMD/DEPAIR	25	2	8	24	54	89	199
TRG SCHOOL	25	4	14	41	101	186	353
MNT SQN	209	14	48	145	211	411	1027
FLY SQN	156	12	40	108	234	391	861
ACFT DEPOT	88	9	31	97	237	422	922
TOTAL SURVEY	515	14	47	141	299	491	1027*

*1027 Tasks Listed in the Instrument Engineering Job Inventory

Tables 3 and 4 present a summary of the average percent time spent on each duty for the major work areas and total survey samples for the electrical and instrument engineering musterings. Inspection of these tables plus Tables 1 and 2 would seem to confirm the idea that the electrical and instrument engineering musterings are extremely heterogeneous in nature, ie, airmen are involved in a wide variety of work tasks spending relatively little time on any of them. This is indicated by the large number of tasks accounting for group work time and by the relatively small amounts of time devoted to non-supervisory duties. As suggested above, these findings were not unanticipated since RAAF airmen are expected to be skilled in all work concerned with their musterings. Electrical fitters, for example, must be proficient with all electrical systems (including ground support equipment) found on every type of aircraft, marine craft, and road transport in the RAAF, while instrument fitters work with all aircraft instruments, instrument systems, and associated equipment including gun and bombsights, cameras, oxygen systems, and cabin pressurization equipment.

Now let us look at some example data which will appear more familiar to you since it closely parallels results from USAF job analyses. Tables 5 and 6 summarize for the six RAAF enlisted ranks the numbers of tasks accounting for 10 percent, 25 percent, 50 percent, 75 percent, 90 percent, and 100 percent of the work time for the electrical and instrument engineering musterings. (AC - aircraftman, LAC - leading aircraftman, CPL - corporal, SGT - sergeant, FSGT - flight sergeant, and WOFF - warrant officer).

/TABLE 3

TABLE 3

AVERAGE PERCENT TIME SPENT ON DUTIES BY THE ELECTRICAL ENGINEERING MUSTERINGS BY MAJOR WORK AREA, AND THE TOTAL SAMPLE (N=496)

DUTY		CMD/DEPAIR	TRG SCHOOL	MNTSQN	ACFT DEPOT	FLYSQN	TOTAL SAMPLE
A.	Planning and Organizing	40.03	11.86	8.41	10.11	6.40	9.65
B.	Directing and Implementing	20.64	9.03	7.06	8.91	4.97	8.09
C.	Inspecting and Evaluating	13.49	5.94	6.74	7.36	4.40	6.22
D.	Reporting and Recording	14.37	3.48	3.79	4.22	3.06	4.10
E.	Training	11.47	58.35	3.85	4.80	3.23	5.94
F.	Servicing Aircraft Systems and Components	0.00	.23	10.95	12.70	13.19	10.80
G.	Servicing Calibration, Test, and Simulator Equipment	0.00	.06	1.16	.53	.73	.89
H.	Servicing Ground Electrical Systems and Equipment	0.00	.60	3.47	3.90	1.78	3.06
I.	Miscellaneous Servicing Activities	0.00	1.22	2.74	2.51	3.72	2.82
J.	Inspecting or Testing Aircraft Systems	0.00	1.18	11.08	8.52	17.38	11.63
K.	Inspecting or Testing Electrical Systems and Equipment	0.00	2.12	5.16	3.86	2.84	4.01
L.	Inspecting or Testing Simulator and Test Equipment	0.00	.14	.58	.31	.09	.34
M.	Troubleshooting Aircraft Systems	0.00	1.17	6.83	5.70	11.48	7.13

DUTY		CMD/DEPAIR	TRG SCHOOL	MNTSQN	ACFT DEPOT	FLYSQN	TOTAL SAMPLE
N.	Troubleshooting Calibration, Test, and Simulator Equipment	0.00	.34	1.21	.64	.39	.91
O.	Troubleshooting Ground Electrical Systems and Equipment	0.00	1.07	2.52	3.30	1.12	2.27
P.	Repairing or Adjusting Aircraft Systems Components	0.00	0.00	6.20	7.04	8.23	6.34
Q.	Repairing or Adjusting Calibration, Test, and Simulation Equipment	0.00	.24	.87	.54	.26	.37
R.	Repairing or Adjusting Ground Electrical Systems and Equipment	0.00	2.08	5.18	5.75	2.54	4.24
S.	Installing or Replacing Aircraft Systems Components	0.00	.45	5.59	3.60	10.74	6.35
T.	Calibrating Measurement, Calibration, and Test Equipment	0.00	.41	.51	.19	.15	.46
U.	Bench Checking Aircraft Systems Components	0.00	0.00	5.35	5.17	3.16	4.18
V.	Bench Checking Calibration, Test, and Simulator Equipment	0.00	.03	.75	.34	.14	.20

TABLE 4

AVERAGE PERCENT TIME SPENT ON DUTIES BY THE INSTRUMENT ENGINEERING MUSTERINGS BY MAJOR WORK AREA, AND THE TOTAL SAMPLE (N=515)

DUTY		CMD/DEPAIR	TRG SCHOOL	MNTSQN	ACFT DEPOT	FLYSQN	TOTAL SAMPLE
A.	Planning and Organizing	39.84	15.41	7.88	9.76	5.94	9.30
B.	Directing and Implementing	18.56	11.92	6.57	8.74	5.82	7.69
C.	Inspecting and Evaluating	22.43	10.84	5.86	8.32	5.24	7.36
D.	Reporting and Recording	11.38	3.38	5.03	7.01	3.56	5.25
E.	Training	6.93	38.74	4.54	4.51	3.79	5.87
F.	Servicing Aircraft Systems and Components	00	1.51	9.88	9.76	10.05	8.96
G.	Servicing Calibration, Test, and Simulator Equipment	00	.96	2.14	2.16	.85	1.74
H.	Miscellaneous Servicing Activities	00	3.04	3.97	6.41	5.75	4.63
K.	Inspecting or Testing Aircraft Systems and Components	00	4.81	12.77	10.79	20.07	13.70
L.	Inspecting or Testing Simulator and Test Equipment	00	.71	1.92	.57	.79	1.32
M.	Troubleshooting Aircraft Systems and Components	.86	1.81	9.25	6.44	12.50	8.81
N.	Troubleshooting Aircraft Systems and Components	00	.70	1.45	2.82	.71	1.93

DUTY		CMD/DEPAIR	TRG SCHOOL	MNTSQN	ACFT DEPOT	FLYSQN	TOTAL SAMPLE
P.	Repairing or Adjusting Aircraft Systems and Components	00	1.17	8.02	5.91	7.18	6.71
Q.	Repairing or Adjusting Calibration, Test and Simulator Equipment	00	1.09	2.20	1.68	.87	1.56
S.	Installing or Replacing Equipment and System Components	00	.78	8.35	4.01	10.99	7.61
T.	Calibrating Measurement, Calibration and Test Equipment	00	.26	.77	3.17	.63	.62
U.	Bench Checking Aircraft Systems and Components	00	2.41	7.28	4.92	4.45	5.50
V.	Bench Checking Calibration, Test and Simulator Equipment	00	.46	2.12	3.02	.81	1.44

Letters I, J, O and R were not used in this survey to designate duties.

NUMBER OF TASKS ACCOUNTING FOR WORK TIME OF
ELECTRICAL ENGINEERING MUSTERINGS BY RANK

		Number of Tasks Accounting For					
Group	N	10% of Time	25% of Time	50% of Time	75% of Time	90% of Time	100% of Time
AC	64	8	26	75	172	288	500
LAC	173	15	45	118	251	424	934
CPL	116	13	44	125	271	442	1109
SGT	90	9	28	83	196	364	1038
FSGT	39	7	23	65	139	261	911
WOFF	16	7	21	55	104	162	461

TABLE 6

NUMBER OF TASKS ACCOUNTING FOR WORK TIME OF
INSTRUMENT ENGINEERING MUSTERINGS BY RANK

		Number of Tasks Accounting For					
Group	N	10% of Time	25% of Time	50% of Time	75% of Time	90% of Time	100% of Time
AC	37	6	18	57	129	210	482
LAC	210	11	35	106	225	401	990
CPL	116	12	43	135	301	517	1022
SGT	95	8	28	85	194	356	856
FSGT	37	9	26	70	140	235	778
WOFF	20	9	24	59	112	174	402

Low values in these columns indicate that group work time is concentrated on a few tasks while high values show that group work time is spread over a large number of tasks. In general, the number of tasks performed increases by rank up through CPL, the journeyman level worker, and then decreases as airmen assume more and more supervisory responsibilities and fewer fitter duties.

The expected area of major difference between ranks would be, of course, the amount of supervision performed. The greater the rank, the greater the percentage of airmen performing supervisory tasks, and greater would be the amount of time spent on supervision. Conversely, as rank increases, the amount of time and the percentage performing tasks outside of supervision would become less significant factors in the job description of airmen. When a comparison of ranks within the electrical and instrument

/engineering

engineering musterings is made, the above described phenomenon does, in fact, occur. Tables 7 and 8 indicate the average percent time spent by the six enlisted ranks across the two musterings being considered for each of the duties in the job inventory.

In certain instances it is important to know in what respects the work performed by groups is dissimilar. Such information may clarify distinctions by ranks or between groups with differing experience or training. A computer program, the group difference description, has been devised to compute differences, in terms of members performing and in terms of average time spent on each task, between any two groups defined by background variables (Morsh and Christal, 1966; Christal, 1971). All tasks in the inventory are ordered according to the magnitude of the differences in percent performing between the two groups compared. The printout begins with the greatest positive difference, proceeds through zero difference, and ends with the greatest negative difference.

After careful review of Table 7 and 8 information, it was concluded that the six ranks could be combined into two meaningful clusters - ACs, LACs, CPLs into a 'worker' group and SGTs, FSGTs, and WOFFs into a 'supervisor' group. Group difference descriptions were then computed. Tables 9 and 10 provide lists of the 24 tasks that most significantly discriminated between the workers and supervisors for the electrical and instrument engineering musterings, and the differences between those percentages for each task. The first dozen tasks performed by 'workers' in each mustering are in the servicing, repairing and adjusting, and installing and replacing duties. Supervisor tasks, on the other hand, fall into the planning and organizing, directing and implementing, and inspecting and evaluating duties. These data lead to the somewhat straightforward interpretation that workers are engaged in the technical aspect of the musterings while the senior NCOs have involved themselves in technical management and supervision. These are, of course, not particularly startling results, but they do provide additional evidence as to the validity of the survey findings.

RAAF Occupational Research Projects

Now let us conclude by very briefly discussing some of our on-going research in support of the occupational analysis program.

One advantage of occupational analysis almost always mentioned in the literature is its use in designing training systems. Yet despite the widespread acknowledgement of the efficacy of selecting instructional objectives from up-to-date job information, only a few researchers (Smith, 1964; Smith, 1966; Ammerman, 1966; Ammerman & Melching, 1966; and Hunter, Lyons, MacCaslin, Smith & Wagner, 1969) have applied themselves to the actual task of developing procedures for translating job data into forms specifically designed to help individuals responsible for curriculum definition. Using Christal's (1970) ideas as our basic guide, we are presently collecting data on five of the eight factors (probability of emergency performance, probability of inadequate performance, consequences of inadequate performance, skill perishability, and task difficulty) he suggests as possible criteria for selecting tasks for inclusion in training courses. Upon completion of our analyses, we hope to develop an algorithm which can be used by training managers to convert occupational information into training objectives and course content. Long range plans call for the development of computer programs which will print out training curricula and trade standards as a routine part of occupational analysis.

/TABLE 7

TABLE 7

AVERAGE PERCENT TIME SPENT ON DUTIES BY THE ELECTRICAL ENGINEERING MUSTERINGS BY RANK

DUTY		AC	LAC	CPL	SGT	FSGT	WOFF
A.	Planning and Organizing	.98	1.60	10.92	19.25	25.25	32.29
B.	Directing and Implementing	.11	1.11	9.90	15.63	16.97	18.57
C.	Inspecting and Evaluating	.04	1.10	6.59	13.56	14.93	23.23
D.	Reporting and Recording	2.87	2.34	4.27	5.62	7.48	7.93
E.	Training	.31	1.59	4.85	14.58	17.37	11.23
F.	Servicing Aircraft Systems and Components	18.75	16.63	10.64	3.18	.63	00
G.	Servicing Calibration, Test, and Simulator Equipment	3.79	1.29	.80	.49	.65	00
H.	Servicing Ground Electrical Systems and Equipment	4.06	4.50	2.89	1.15	.41	00
I.	Miscellaneous Servicing Activities	6.04	3.71	2.46	.95	.18	00
J.	Inspecting or Testing Aircraft Systems	17.81	13.31	10.52	7.81	6.64	3.31
K.	Inspecting or Testing Electrical Systems and Equipment	5.13	4.91	3.87	2.58	1.98	1.69
L.	Inspecting or Testing Simulator and Test Equipment	.11	.45	.28	.32	.32	.06

DUTY	AC	LAC	CPL	SGT	FSGT	WOFF
M. Troubleshooting Aircraft Systems	7.54	9.07	6.98	6.12	2.28	.60
N. Troubleshooting Calibration, Test, and Simulator Equipment	.18	1.14	1.60	.81	.62	.16
O. Troubleshooting Ground Electrical Systems and Equipment	2.10	3.16	1.93	1.27	.59	.21
P. Repairing or Adjusting Aircraft Systems Components	9.46	10.13	6.17	1.44	.48	.24
Q. Repairing or Adjusting Calibration, Test, and Simulation Equipment	.29	.86	.45	.38	.70	00
R. Repairing or Adjusting Ground Electrical Systems and Equipment	6.56	6.12	4.00	1.29	.55	.19
S. Installing or Replacing Aircraft Systems Components	10.29	9.25	6.66	1.54	.33	00
T. Calibrating Measurement, Calibration, and Test Equipment	.06	.20	.30	.52	.65	00
U. Bench Checking Aircraft Systems Components	3.40	7.11	3.67	1.08	.26	.29
V. Bench Checking Calibration, Test, and Simulator Equipment	.12	.42	.25	.43	.73	00

TABLE 8

AVERAGE PERCENT TIME SPENT ON DUTIES BY THE INSTRUMENT ENGINEERING MUSTERINGS BY RANK

DUTY		AC	LAC	CPL	SGT	FSGT	WOFF
A.	Planning and Organizing	.30	1.79	8.14	19.74	27.17	28.32
B.	Directing and Implementing	1.49	2.22	8.92	15.49	15.96	18.43
C.	Inspecting and Evaluating	1.17	2.15	6.27	14.58	17.04	24.18
D.	Reporting and Recording	5.28	4.59	4.89	5.67	7.99	8.78
E.	Training	.55	1.70	4.72	14.64	12.97	12.14
F.	Servicing Aircraft Systems and Components	15.42	13.06	9.60	2.54	1.09	.35
G.	Servicing Calibration, Test, and Simulator Equipment	.53	2.06	2.35	1.30	.77	00
H.	Miscellaneous Servicing Activities	6.06	6.16	3.84	1.35	1.57	.16
K.	Inspecting or Testing Aircraft Systems and Components	21.04	16.93	12.97	7.99	6.58	4.43
L.	Inspecting or Testing Simulator and Test Equipment	.29	1.49	1.17	1.32	.37	.82
M.	Troubleshooting Aircraft Systems and Components	8.71	11.52	9.24	6.69	4.30	2.21
N.	Troubleshooting Calibration, Test and Simulator Equipment	.49	1.79	1.96	1.61	.66	.07

DUTY		AC	LAC	CPL	SGT	FSGT	WOFF
P.	Repairing or Adjusting Aircraft Systems and Components	12.19	9.95	6.30	1.88	.67	.01
Q.	Repairing or Adjusting Calibration, Test and Simulator Equipment	.60	2.02	2.31	.99	.26	00
S.	Installing or Replacing Equipment and System Components	10.09	11.29	8.21	1.81	.16	00
T.	Calibrating Measurement, Calibration and Test Equipment	.68	1.61	1.79	.20	.82	00
U.	Bench Checking Aircraft Systems and Components	13.47	7.32	4.70	1.50	.79	.08
V.	Bench Checking Calibration, Test and Simulator Equipment	1.64	2.35	2.62	.70	.83	.02

Letters I, J, O and R were not used in this survey to designate duties.

TABLE 9

MOST DISCRIMINATING TASKS - ELECTRICAL ENGINEERING MUSTERINGS
ACs, LACs, CPLs - SGTs, FSGTs, WOFFs

	Percent of Members Performing in Group		
ACs, LACs, CPLs	ACs, LACs, CPLs	SGTs, FSGTs, WOFFs	Difference
Service DC Power Generation Systems	57.22	13.10	44.12
Service AC and DC Power Control and Distribution Systems	56.94	15.17	41.77
Service Aircraft Interior Lighting Systems	56.09	14.48	41.61
Service Aircraft Exterior Lighting Systems	55.52	14.48	41.04
Install or Replace Aircraft Interior Lighting System Components	50.42	11.03	39.39
Install or Replace Aircraft Exterior Lighting System Components	50.14	11.03	39.11
Install or Replace AC and DC Power Control and Distribution System Components	45.33	7.59	37.74
Repair or Adjust Aircraft Interior Lighting System Components	48.16	11.72	36.44
Service AC Power Generation Systems	50.42	14.48	35.94
Repair or Adjust Aircraft Exterior Lighting System Components	47.59	11.72	35.87
Service Battery Control and Distribution Systems	48.16	12.41	35.75
Install or Replace Aircraft Wiring Components	43.91	8.28	35.63
SGTs, FSGTs, WOFFs			
Assist Officer-in-Charge in Establishing Organizational Policy	7.08	58.62	-51.54
Evaluate Compliance with Established Work Standards	10.20	62.07	-51.87
Enforce Fire and Safety Regulations	10.20	62.07	-51.87

SGTs, FSGTs, WOFFs	Percent of Members Performing in Group		
	ACs, LACs, CPLs	SGTs, FSGTs, WOFFs	Difference
Supervise the Preparation and Maintenance of Reports and Records	8.22	60.69	-52.47
Develop or Revise the Organization of the Section	10.76	63.45	-52.69
Allot Section Work Priorities	15.30	68.28	-52.98
Control Flow of Work	17.28	70.34	-53.06
Ensure Display of Appropriate Notices	9.35	62.76	-53.41
Evaluate Competency of Airmen to Perform Assigned Task	15.86	70.34	-54.48
Supervise Section Security	13.03	67.59	-54.56
Counsel Personnel Regarding Personnel Problems	5.67	64.83	-59.16
Evaluate Individuals for Promotion or Upgrading	8.78	75.17	-66.39

TABLE 10

MOST DISCRIMINATING TASKS - INSTRUMENT ENGINEERING MUSTERINGS
ACs, LACs, CPLs - SGTs, FSGTs, WOFFs

	Percent of Members Performing in Group		
ACs, LACs, CPLs	ACs, LACs, CPLs	SGTs, FSGTs, WOFFs	Difference
Service Aircraft Instrument Gauges and Indicators	56.20	16.34	39.86
Install or Replace Altimeters and Rate of Climb Indicators	45.45	6.54	38.91
Install or Replace Airspeed Indicators	44.90	6.54	38.36
Marshal Aircraft	43.80	6.54	37.26
Prepare and Package Equipment for Transit or Storage	46.56	9.80	36.76
Install or Replace Aircraft Fuel Contents Gauges	43.53	7.19	36.34
Service Aircraft Liquid and Gaseous Oxygen Systems	44.08	9.80	34.28
Refuel Aircraft	38.02	3.92	34.10
Service Aircraft Pitot Static Systems	43.80	9.80	34.00
Install or Replace Aircraft Clocks	40.50	6.54	33.96
Clean Aircraft	35.26	2.61	32.65
Install or Replace Aircraft Engine Tachometers	37.19	4.58	32.61
Install or Replace Breathing Oxygen System Components	40.22	7.84	32.38
SGTs, FSGTs, WOFFs			
Recommend Variations to Technical Maintenance Plans	4.68	58.17	-53.49
Develop and Improve Work Methods	22.08	75.82	-53.74
Evaluate Compliance with Established Work Standards	9.37	64.05	-54.68

SGTs, FSGTs, WOFFs	Percent of Members Performing in Group		
	ACs, LACs, CPLs	SGTs, FSGTs, WOFFs	Difference
Ensure Display of Appropriate Notices	4.68	59.48	-54.80
Control Flow of Work	14.05	69.28	-55.23
Enforce Fire and Safety Regulations	10.47	66.01	-55.54
Allot Section Work Priorities	19.56	75.16	-55.60
Examine Technical Publications for Completeness and Accuracy	13.77	69.93	-56.16
Counsel Personnel Regarding Personal Problems	4.68	63.40	-58.72
Assign Specific Work to Individuals	26.17	84.97	-58.80
Evaluate Competency of Airmen to Perform Assigned Tasks	15.98	75.16	-59.18
Evaluate Individuals for Promotion or Upgrading	9.09	72.55	-63.46

A second project on which we are currently working concerns job satisfaction within the RAAF. As more and more emphasis is placed on improved personnel utilization within an environment of defence strength and budget reductions, job satisfaction has begun to take on an increasing importance within the military services. In 1966, the US Air Force initiated a program to evaluate the impact of work assigned on job satisfaction, felt utilization of talent, and career decisions (Gould & Christal, 1971; Gould, 1972; and Christal, 1972). Respondents to job inventories use a 7-point scale to indicate their job interest ranging from 'extremely dull' to 'extremely interesting' and the extent to which their jobs use their talents and training with responses from 'not at all' to 'perfectly'.

Results to date appear very promising and suggest a powerful methodology for studying the attitudes of airmen toward their work, and for identifying possible work changes within musterings which may improve job satisfaction and thus work performance (Stacy, 1972). Consequently, we are now capturing job satisfaction data as a regular part of our occupational survey procedures. Unfortunately, we do not yet have any results to present to you; however, it is hoped that by next year's conference we will be able to discuss job satisfaction in the RAAF in some detail.

Summary

The future of occupational analysis in the Royal Australian Air Force looks promising. The feasibility of the job inventory approach has been demonstrated, and we are constantly striving to develop innovations and advances in our inventory construction and survey administration techniques. We plan an expansion of our occupational survey schedule, and this, combined with our anticipated research results, should provide the RAAF with job information which will lead to improvements in the management of personnel and training programs.

REFERENCES

Ammerman, H.L. *Development of procedures for deriving training objectives for junior officer jobs.* HUMRRO TR-66-3. Alexandria, VA: Human Resources Research Office, May 1966.

Ammerman, H.L. and Melching, W.H. *The derivation, analysis and classification of instructional objectives.* HUMRRO TR-66-4. Alexandria, VA: Human Resources Research Office, May 1966.

Archer, W.B. *Computation of group job descriptions from occupational survey data.* PRL-TR-66-12. Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division, December 1966.

Christal, R.E. Implications of Air Force occupational research for curriculum design. Paper presented to the Minnesota Research Co-ordinating Unit for Vocational Education Seminar: Process and Techniques of Vocational Curriculum Development. University of Minnesota, November 1970.

Christal, R.E. Combining occupational information with personnel data to resolve manpower management problems. Paper presented to the NATO Conference on Manpower Planning Models, Cambridge, England, September 1971.

Christal, R.E. New directions in the Air Force occupational research program. Paper presented to the 80th Annual Convention of the American Psychological Association, Honolulu, September 1972.

Gould, R.B. *Reported job interest and perceived utilization of talents and training by airmen in 97 career ladders.* AFHRL-TR-72-7. Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory, January 1972.

Gould, R.B. and Christal, R.E. Enlisted job satisfaction in the Air Force: A study at the task level. Paper presented to the 13th Annual Conference of the Military Testing Association, Washington DC, September 1971.

Hunter, H.G., Lyons, J.D., MacCaslin, E.F., Smith, R.G., and Wagner, H. *The process of developing and improving course content for military technical training.* HUMRRO TR-69-9. Alexandria, VA: Human Resources Research Office, May 1969.

Morsh, J.E. Job analysis in the United Stated Air Force. *Personnel Psychology*, 1964, 37, 7-17.

Morsh, J.E. Collecting, analysing, and reporting information describing jobs in the United States Air Force. Paper presented to the 77th Annual Convention of the American Psychological Association, Washington DC, September 1969.

Morsh, J.E. and Christal, R.E. *Impact of the computer on job analysis in the United States Air Force.* PRL-TR-66-19. Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division, October 1966.

Morsh, J.E., Madden, J.M., and Christal, R.E. *Job analysis in the United States Air Force.* WADD-TR-61-113. Lackland AFB, TX: Personnel Laboratory, Wright Air Development Division, February 1961.

Sellman, W.S. and Crow, R.R. *Occupational survey of the electrical engineering musterings.* GT3-JA-72-3. Melbourne, Australia: HQ Support Command, Royal Australian Air Force, December 1972.

Sellman, W.S. and Crow, R.R. *Occupational survey of the instrument engineering musterings.* GT3-JA-73-1. Melbourne, Australia: HQ Support Command, Royal Australian Air Force, April 1973.

Smith, R.G. *The development of training objectives.* HUMRRO Bulletin II. Alexandria, VA: Human Resources Research Office, May 1964.

Smith, R.G. *The design of instructional systems.* HUMRRO TR-66-18. Alexandria, VA: Human Resources Research Office, November 1966.

Stacy, W.J. *Felt utilization of talent and training in two civil engineering career ladders.* AFHRL-TR-72-XX. Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory (In Press).

COLLECTION OF PERFORMANCE DATA AT THE JOB TASK LEVEL*

CLIFFORD P. HAHN

I want to briefly describe a procedure AIR is using in a current study for the Air Force to develop task-level job performance criteria. The basic concept is one of multi-variate indices to measure job performance. A major distinguishing characteristic of the study is the use of 40 or more task-level performance indices per job as opposed to a few global measures.

Perhaps it should be pointed out that a fundamental reason for undertaking the study was not to come up with measures of performance for the sake of determining individual worth, but rather to derive measures that could be used to validate various portions of existing selection, classification, assignment, and training procedures. The airmen included in the study had initially been selected with aptitude tests, had been assigned to gross occupational areas through the use of clustered aptitude scores in a classification process, and had been trained through a combination of technical schools and on-the-job programs. The job performance measured during this study of journeyman-level airmen is thus partly a product of their inherent personal capabilities and partly a product of the selection, classification, assignment, and training systems they went through. The basic purpose was to use the indices developed to validate the system outputs. Since individual job performance represented a direct index of that output, it is hoped that it can be used for these purposes.

As indicated earlier, the basic concept is one of multiple criteria for evaluating a total job, a major job segment, or with appropriate adaptations, major dimensions of job performance. If the jobs themselves are made up of a series of separate identifiable tasks, it seems appropriate to measure success on those jobs by measuring success on the discrete tasks which make up the jobs. It is assumed that within any given occupational area there are an identifiable number of job tasks which can be used to

*Paper for presentation at the 1973 Military Testing Association Convention, San Antonio, Texas on 28 October - 2 November 1973.

describe all jobs within that area. No single job will involve all of the tasks, yet all jobs within the occupational area can be described by some combination of the tasks within the occupational area. If procedures exist for obtaining performance indices for all tasks, then any given individual's job performance can be appraised only on those tasks included in his job. The Air Force system of task analysis certainly provides the necessary information on a descriptive basis, and we are attempting to develop relevant performance information at the same level. Existing techniques for clustering tasks into duty positions and occupational fields should be equally appropriate for combining task performance indices into clustered performance indices either on the basis of job duties, total jobs, or job dimensions. If this basic concept of performance evaluation can be effectively implemented, it has tremendous potential for validation studies of all kinds.

The most comprehensive implementation of a task level model would include all possible tasks and this would be theoretically possible. In most practical situations, however, and certainly in our study, it was necessary to pare the total task lists down to a manageable number. A number of different approaches are possible. We used the collective judgments of experienced first and second line supervisors in an iterative process involving first groups of two or three, then ten, and then approximately fifty. How many tasks can be included depends largely upon the resources available for implementing the study or operational procedure.

Implementation of the task level model in its most comprehensive form would also call for the measurement of all relevant dimensions of performance for each job task selected. This too, while theoretically possible, would be difficult in most practical situations. It is important however, to recognize that there are several different dimensions of task performance for most tasks. Included are such dimensions as quality, speed, materials and facilities used, amount of outside help needed, self-initiation and others. Again, we used groups of experienced first and

second line supervisors to judge which of the dimensions for a given task were most critical. Critical in the sense that they provided the basis for discriminating between good and poor performers. This process in itself was illuminating. For certain task dimensions, the supervisors were able to identify only two groups, an acceptable performance group and an unacceptable group. For other task dimensions, they were able to identify a high, medium, and low group, and for a relatively few task dimensions they could identify a super group, an average expected group, a minimally acceptable group, and an unacceptable group. Quite obviously, in working with the task level approach, the scalability of task dimensions must be investigated. For practical reasons, we ended up using a standard scale for all task dimensions, but it will be interesting to see if the judgmentally identified groups will be supported by the empirical data.

The process for generating performance indices on the selected task dimensions included self-ratings, supervisor ratings and peer ratings. While the liabilities of ratings as performance indices were fully recognized, in this case the requirement for developing a system that would work for large numbers of people scattered across the world within feasible cost parameters almost dictated the use of ratings. One of the questions to be answered by our effort is whether or not ratings at the task level diminish or eliminate some of the previously identified problems of ratings as measures of job performance.

Since we have performance ratings from job incumbents, supervisors, and peers we will be able to determine the degree to which the three groups agree on what tasks are included in the job of the person being evaluated. Unless there is good agreement between the groups, rated performance is obviously based on different rather than identical concepts of the job. In the typical global rating situation one has no data at all on this aspect of the performance rating process. Table 1 shows some data from a preliminary analysis of 46 cases in the Communications Center Operations. There is reasonably good agreement although far from perfect and it varies from task to task. In more cases both the supervisors and peers underinclude, although there are most exceptions to this for the supervisors than the peers.

Analyses of the extent of agreement between self-ratings and ratings of supervisors and peers as well as between two supervisors or two peers for the same incumbent and between supervisors and peers will be made. Depending upon the relationships obtained, it may be possible to identify certain tasks or groups of tasks which can best be rated only by supervisors, or only by peers, or equally by both. There will probably be tasks for which neither supervisor or peer ratings are really appropriate. These should obviously be specifically excluded from the judgmental base for making performance ratings. Other tasks might be excluded on the theoretical basis that they provide no performance variance and therefore cannot help discriminate between different levels of performance. If one were using the performance ratings as a measure of personal worth, this would not be defensible since such tasks are a part of the total performance picture. If the performance ratings are used to judge the value of personnel system outputs as they are intended to in this case, the elimination of such non-variance tasks is both defensible and desirable.

Table 2 shows some data from a preliminary analysis of performance ratings for 47 cases in the Communications Center Operations career field. As you can see, the range of correlations is considerable. There appears to be more agreement among supervisors and peers than among incumbents and either group, with one exception.

Another aspect of the task level approach as we are attempting to develop it is our attempt to tease out from the levels of job performance obtained that portion which is due to skill and ability factors and that portion due to motivational factors. In order to obtain data bearing on this question, incumbents were asked to rate both the intensity and direction of six factors for each task. These factors included the nature of the actual work, the importance of the task, the match between task difficulty and incumbent capabilities, the opportunity for praise and recognition, the amount and kind of supervision, and personal relations. Incumbents were also provided the opportunity for write-in factors of their own choosing.

AD-A040 773 AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TEX F/G 5/9
PROCEEDINGS, ANNUAL CONFERENCE OF THE MILITARY TESTING ASSOCIAT--ETC(U)
NOV 73

UNCLASSIFIED NL

8 OF 9
ADA
040773



Table 3 shows some data from a preliminary analysis of the motivational ratings for 43 cases from the Communications Center Operations career field.

A review of Table 3 reveals the relatively small per cent of all possible tasks ranked as being done and having an appreciable effect on motivational level for a given duty position. This would tend to indicate that there may be a great deal of masking of results when overall motivational ratings across fairly broad occupational groups are obtained. It appears that self-judged importance of the task is the predominant motivational factor at the task level, and that it is largely positive in nature. On the other hand, the opportunity for praise and recognition appeared to have much less potency and it was about as likely to be positive as negative in nature. Four of the six factors tended to be largely positive in nature.

Supervisors and peers, after having ranked the performance level of an incumbent on a task dimension, were asked to make a judgment regarding what it would take for the incumbent to reach his full performance potential on that task--either the acquisition of additional skills and knowledge, the raising of his motivational level, or combinations of both. A five-point scale was used for this. Analyses of these data should give an indication of the types of tasks for which performance levels are dictated largely by skill and ability factors, those which are largely dictated by desire to do a good job, and those which are an admixture of both. Skill determined tasks are normally thought of as somewhat unique, while motivationally determined tasks could probably be expected to be task-general, but person-unique. Whether or not these expectations will be shown to be empirically true remains to be seen.

If analyses of the data collected indicate that we have produced the kinds of indices that we had hoped we would, the major impact on the potential for better utilization of human resources will derive from the fact that there will be an abundance of data for many different validity studies. Each segment of the personnel management system, e.g., selection, classification, assignment, and training, can be differentially subjected to more meaningful validation studies.

Table 1

Agreement Between Incumbents, Supervisors, and Peers on Tasks Included in Incumbents Job
(N-46 - 29150 AFSC)

Task Dimension	% Incumbents Inclusion	% Supervisors Inclusion	% Peer Inclusion	Supervisors Inclusion as % of Incumbent Inclusion	Peer Inclusion as % of Incumbent Inclusion
1	.80	.73	.64	.91	.80
2A	.50	.48	.38	.96	.76
2B	.52	.50	.41	.96	.79
3	.70	.71	.59	1.01	.84
4A	.34	.36	.38	1.06	1.12
4B	.34	.34	.37	1.00	1.09
5	.67	.71	.58	1.06	.87
6	.63	.70	.55	1.11	.87
7	.76	.68	.55	.89	.72
8A	.72	.63	.56	.88	.78
8B	.85	.70	.56	.82	.66
8C	.72	.61	.56	.85	.78
9	.78	.77	.65	.99	.83
10	.67	.64	.65	.96	.97
11A	.76	.70	.59	.92	.78
11B	.70	.66	.55	.94	.79
12	.57	.50	.41	.88	.72
13	.78	.71	.60	.91	.77
14	.63	.70	.54	1.11	.86
15	.74	.68	.51	.92	.69
16	.74	.70	.55	.95	.74
17	.67	.61	.50	.91	.75
18	.67	.50	.47	.75	.70
19	.61	.45	.44	.74	.72
20	.57	.59	.47	1.04	.82
21	.80	.64	.55	.80	.69
22	.76	.73	.49	.96	.64
23A	.61	.52	.41	.85	.67
23B	.48	.43	.41	.90	.85
24	.52	.50	.44	.96	.85
25	.43	.14	.37	.33	.86
26	.37	.11	.29	.30	.78
27	.46	.29	.29	.63	.63
28	.43	.29	.28	.67	.65
29	.30	.21	.19	.70	.63
30	.33	.18	.18	.55	.55
31	.41	.27	.27	.66	.66

Table 1 (Continued)

Task Dimension	% Incumbents Inclusion	% Supervisors Inclusion	% Peer Inclusion	Supervisors Inclusion as % of Incumbent Inclusion	Peer Inclusion as % of Incumbent Inclusion
32	.37	.29	.26	.78	.70
33	.41	.25	.24	.61	.59
34	.48	.45	.40	.94	.83
35	.54	.55	.35	1.02	.65
36	.50	.45	.40	.90	.80
37	.52	.36	.31	.69	.60
38	.46	.38	.35	.83	.76
39	.48	.45	.40	.94	.83
40	.57	.46	.38	.81	.67
41	.61	.50	.42	.82	.69
42	.43	.32	.24	.74	.56
43A	.70	.64	.45	.91	.64
43B	.67	.61	.42	.91	.63
44	.67	.57	.36	.85	.54

Table 2

Correlations Between Performance Ratings
by Different Types of Raters
29150 - Communications Center Operations
(Based on 47 Cases with Ratings Aggregated Across all Tasks' Rates)

	Supervisor 1	Supervisor 2	Peer 1	Peer 2
Incumbent	.29 N= 47 n=999	.93 N= 10 n=235	.27 N= 43 n=805	.09 N= 32 n=603
Supervisor 1		.50 N= 11 n=234	.34 N= 43 n=792	.34 N= 31 n=568
Supervisor 2			.27 N= 9 n=209	
Peer 1				.51 N= 32 n=482

Table 3

Distribution of Incumbents' Motivational Rankings
(Based on 43 cases each having potential rankings for 44 tasks)
(Most Important = 1 Least Important = 6)

Motivational Factor	Mean Ranking All Tasks Included	Per Cent of all Tasks Ranked	Mean Number of Tasks Ranked	Mean Ranking of Tasks Ranked	Direction of Motivational Influences - Per Cent of all Tasks Ranked		
					Positive	Negative	P/N
Importance of Task	.58	42	793	1.39	81	8	11
Nature of Actual Work	.57	30	570	1.90	74	15	11
Match Between Task Difficulty and Incumbent Capability	.36	15	274	2.50	72	19	9
Amount and Quality of Supervision	.37	15	282	2.53	72	14	14
Personal Relations	.29	10	197	2.85	55	6	39
Opportunity for Praise and Recognition	.29	9	170	3.27	40	33	27

Table 4

Mean Motivational Ranks by Task and Factor

Task Number	Importance of Task	Nature of Work	Match Between Difficulty and Capability	Amount and Quality and Supervision	Personal Relations	Opportunities for Praise and Recognition
1	1.30	1.95	2.50	2.71	2.50	3.40
2	1.29	2.70	2.50	1.90	2.50	3.33
3	1.29	1.93	2.71	1.80	2.67	2.50
4	1.44	1.75	3.33	2.40	3.50	4.00
5	1.22	2.07	2.60	3.00	3.38	3.00
6	1.20	1.71	3.00	3.50	2.33	4.00
7	1.31	1.94	1.88	2.40	1.67	2.00
8	1.46	1.81	2.08	3.20	3.00	3.75
17	1.25	1.77	2.83	2.29	4.00	4.00
18	1.47	1.79	3.00	2.17	3.67	2.67
25	1.29	2.08	3.29	2.13	3.25	3.67
26	1.33	1.82	2.33	4.00	1.67	1.50
27	1.42	2.43	2.50	2.00	2.40	3.14
28	1.73	1.88	2.67	2.00	1.25	2.67
33	1.63	1.63	2.75	1.60	3.67	3.33
34	1.29	2.00	2.50	3.44	3.00	3.33
35	1.40	2.17	2.63	2.22	2.40	2.67
40	1.29	1.60	1.67	2.00	2.33	1.00
41	1.37	1.71	2.00	2.86	3.67	3.50
44	1.26	1.83	3.00	2.00	4.00	5.00
Mean Across Tasks	1.36	1.93	2.59	2.48	2.84	3.12
Range	1.20-1.73	1.60-2.70	1.67-3.33	1.60-3.50	1.25-4.00	1.00-5.00
*S.D.	.13	.27	.44	.65	.80	.93

*S.D. = Standard Deviation

Table 5

Motivational Direction of 'Nature of Work for 20 Tasks'

Task Number	Per Cent of Rankings Made			Number of Ranks Made
	Positive	Negative	P/N	
1	62	10	28	21
2	70	0	30	10
3	60	33	7	15
4	75	25	0	4
5	71	7	22	14
6	83	8	8	12
7	89	11	0	18
8	81	14	5	21
17	54	38	8	13
18	79	14	7	14
25	75	17	8	12
26	64	18	18	11
27	43	57	0	7
28	63	25	12	8
33	38	50	12	8
34	93	0	7	15
35	100	0	0	12
40	40	30	30	10
41	57	21	22	14
44	75	17	8	12
Mean Across Tasks	68.6	19.8	11.6	
Range	38-100	0-57	0-30	
Standard Deviation	17.1	15.7	10.0	

Table 6

Motivational Direction of 'Importance of Task'

Task Number	Per Cent of Rankings Made			Number of Ranks Made
	Positive	Negative	P/N	
1	81	12	7	26
2	79	0	21	14
3	82	12	6	17
4	78	0	22	9
5	94	0	6	18
6	80	10	10	20
7	88	4	8	26
8	81	0	19	22
17	88	6	6	16
18	59	35	6	17
25	81	9	10	21
26	83	8	9	12
27	58	17	25	12
28	55	9	36	11
33	75	12	13	8
34	88	12	0	17
35	95	0	5	20
40	82	12	6	17
41	80	16	4	19
44	84	5	11	19
Mean Across Tasks	79.6	9.0	11.5	
Range	55-95	0-35	0-36	
Standard Deviation	10.8	8.2	8.8	

Table 7

Motivational Direction of 'Match Between Task Difficulty and Incumbent Capability'

Task Number	Per Cent of Rankings Made			Number of Ranks Made
	Positive	Negative	P/N	
1	83	17	0	6
2	72	12	13	8
3	86	14	0	7
4	100	0	0	3
5	90	0	10	10
6	75	25	0	4
7	75	25	0	8
8	67	8	25	12
17	83	17	0	6
18	20	60	20	5
25	57	14	29	7
26	50	33	17	6
27	33	50	17	6
28	33	67	0	3
33	50	50	0	4
34	100	0	0	8
35	63	25	12	8
40	33	33	34	3
41	75	25	0	4
44	83	0	17	6
Mean Across Tasks	66.6	23.8	9.7	
Range	33-100	0-67	0-34	
Standard Deviation	23.5	20.1	11.3	

Table 8

Motivational Direction of 'Amount and Quality of Supervision'

Task Number	Per Cent of Rankings Made			Number of Ranks Made
	Positive	Negative	P/N	
1	57	29	14	7
2	80	10	10	10
3	60	40	0	5
4	70	10	20	10
5	67	0	33	9
6	100	0	0	2
7	80	0	20	5
8	60	0	40	5
17	43	43	14	7
18	33	50	17	6
25	63	25	12	8
26	100	0	0	2
27	33	50	17	6
28	40	40	20	5
33	80	20	0	5
34	78	11	11	9
35	78	0	22	9
40	75	25	0	4
41	100	0	0	7
44	100	0	0	8
Mean Across Tasks	69.9	17.7	12.5	
Range	33-100	0-50	0-40	
Standard Deviation	21.6	18.7	11.7	

Table 9

Motivational Direction of 'Personal Relations'

Task Number	Per Cent of Rankings Made			Number of Ranks Made
	Positive	Negative	P/N	
1	33	17	50	6
2	25	0	75	8
3	67	0	33	3
4	67	0	33	6
5	63	12	25	8
6	67	0	33	3
7	83	0	17	6
8	71	0	29	7
17	33	0	67	3
18	33	33	34	6
25	75	0	25	4
26	67	0	33	3
27	60	20	20	5
28	75	0	25	4
33	67	0	33	3
34	50	0	50	4
35	60	0	40	5
40	33	0	67	3
41	67	0	33	3
44	75	0	25	4
Mean Across Tasks	58.6	4.1	37.4	
Range	25-83	0-33	17-75	
Standard Deviation	17.5	9.1	16.3	

Table 10

Motivational Direction of 'Opportunity for Praise and Recognition'

Task Number	Per Cent of Rankings Made			Number of Ranks Made
	Positive	Negative	P/N	
1	40	20	40	5
2	67	0	33	3
3	50	25	25	4
4	33	33	34	3
5	67	33	0	3
6	100	0	0	1
7	50	50	0	2
8	50	25	25	4
17	100	0	0	3
18	33	50	17	6
25	33	33	34	3
26	0	100	0	2
27	0	71	29	7
28	20	60	20	5
33	0	33	67	3
34	67	33	0	3
35	67	33	0	3
40	0	100	0	1
41	0	50	50	4
44	25	0	75	4
Mean Across Tasks	40.1	37.5	22.5	
Range	0-100	0-100	0-75	
Standard Deviation	31.8	29.3	23.3	

COGNITIVE COMPLEXITY: ITS ASSOCIATION WITH SELECTION FOR MILITARY LEADERSHIP ROLES

SQUADRON LEADER BRIAN N PURRY RAF
DEFENCE FELLOW, DEFENCE STUDIES UNIT
31 BUCCLEUCH PLACE, UNIVERSITY OF
EDINBURGH, SCOTLAND

Abstract

The concept of cognitive complexity, its relationship to personality variables, and in particular its association with the performance of leadership functions is briefly reviewed. Four postulates are advanced and examined:

Prevailing contingency theories may, by over-emphasising the situational deteminants of leadership, tend to obscure permanent sources of prominence and influence operating in a wide range of situations.

Intelligence, conventionally measured as a convergent ability, does broadly predict potential occupational status; but

Intelligence, above certain cut-offs and within fairly broad bands, neither predicts excellence, nor does it aid vocational choice, nor (especially) does it distinguish potential leaders; but

Cognitive complexity, and specifically that element which is defined as the ability to differentiate, is a permanently distinguishing characteristic of potential leaders.

The construction of an experimental differentiation test or exercise is described and discussed in detail and it is concluded that objective measures of cognitive complexity may usefully be added to leader selection batteries and appraisal training programmes so as to enhance the utilisation of human resources in conditions of growing demands.

Intentions, Assumptions and Plan

I want, if I may, to talk briefly about intelligence and leadership: not to define nor really to explain either, since far better men than I have stumbled on that rock, but to investigate with you a small part of the association of these two variables and to explore an idea that objective measurement of a particular facet of the former may usefully predict performance in some functional aspects of the latter.

I start with the belief that since the extensive and well documented controversy over the relative merits of clinical and statistical prediction (1), there is, to put it at its mildest, doubt thrown on the reliability and validity of subjective evaluations made in selection. Any objective measure, therefore which usefully correlates with some aspect of leadership performance is to be welcomed.

---

1. See for example Clinical Vs statistical prediction, Meehl 1954, University of Minnesota Press; or Measurement and prediction, clinical and statistical, Sawyer, Psychological Bulletin Vol 66, No 3, Sept 66.

Not that everyone, of course, subscribes to this view of appraisal. In a recent Sunday Times article on selecting a pony for a child, readers were advised to look for large intelligent eyes, whilst the guidance given in this matter of world importance included the following:
(and I quote)

"Kind ponies have kindly eyes, piggy little eyes reveal piggy little natures" (2)

Nor was objectivity always attainable or valued by professionals. You will recall that for Spearman when writing early in his career at the turn of the century, intelligence was just four kinds of mental activity refreshingly simple to conceptualise and even easier to assess. Two of these you will remember were:

"General impression produced upon other people, ie bright, average, or dull and

"Commonsense which was established by asking the oldest pupil to say who had the most 'sharpness and commonsense out of school', and who next, and so on.

And finally, by way of a reliability check, the supreme test, "a similar list was obtained from the rector's wife who also had always lived in this village, but her graduation is unfortunately incomplete and therefore unusable....." (3)

At least the minister's wife had always lived in the village - an experience few personnel selection officers can call on.

There is now of course a multiplicity of intelligence and personality tests, motivation measures, and so on but there is still a need to find simple objective measures directly germane to the identification of leadership potential.

I shall try therefore to describe my own first faltering steps in this direction towards one, look in detail at construction of the instrument, and finally discuss its use in experimental and training situation. I would welcome comment upon this line of investigation, but first let me set the scene with some thoughts on concepts of mind and of leadership which it seems to me are not without relevance.

---

2. Daddy, daddy can I please have a pony, Silver and Bousefield, Sunday Times, 8 Apr 1973.

3. General intelligence objectively determined and measured, Spearman C, American Journal of Psychology, Vol 115, 1904.

Of Minds and Master Minds

The concept of mind has always been something of a puzzle. A mystery somewhat akin to 'soul', it baffles the intellect, evades logic, eludes comprehension. Physiological knowledge of its fundamental parts and their functioning, as yet in comparative infancy, relates little and in but few cases to the maps and schema of psychology. Whether these theoretical constructs be derived by simple intuitive introspection, or from more sophisticated factorial analyses, or indeed from the theory of radio transmitters and receivers, they do not physically fit. The neurological structure of mind bears resemblance neither to Vernon's heirarchy of abilities (4) nor to Guilford's cube (5). Whilst the literature of psychology does strive after a tidy, neat and comprehensible system of classification and order, even the boundaries between its several branches lack precision and are not unambiguous but the line of demarcation between thought and personality in particular, between ability and traits, between cognitive style and intelligence is so blurred and intangible as to be non-existent.

Even intelligence, more often measured, tested, analysed, diagnosed, explained, written about than any other single human attribute, is yet but an operational definition. We know not what it really is. In the linguistic and conceptual chaos of the psychology of personality and mind it is a brave and foolish man who dares suggest he understands. So also is it with the equally baffling matter of leadership style and potential - charisma and capability. But let us try and unravel some of this.

Like Ceasar's Gaul, the long history of leadership studies falls naturally into three parts which in this company hardly need mention. In the first, biography and autobiography, reports, news items, letters, formal and informal accounts of great men were scoured for the personality attributes of charismatic greatness. Personal character qualities were thought to be the defining characteristics of the leader and indeed there still are appraisal systems in which this somewhat impracticable premise plays a dominant part. In the second phase, emphasis shifted from what the leader is to what he does. Small group theory held sway and since any sociologist half worthy of the name can always rustle up a few subjects for a quick research, there must now be a billion billion studies of the factors affecting the emergence of a leader in a small informal group. Two-person groups, triads, fours and tens, billions of them, and since inevitably many of the findings in these functional studies

---

4. The structure of human abilities, Vernon P E, Methuen 1950.

5. Three faces of intellect, Guilford J P, American Psychologist Vol 14, 1959 and others by the same author.

contradicted the others there developed the third phase which may be called, "it all depends on", or, "leadership is situationally determined", as for example in J M Barrie's tale of the Admirable Crichton, or as was more formally stated by Warr:

> "This leads to the paradox that psychology (which is usually seen as a study of people) must turn more to study environments and situations if its developmental momentum is not to be lost". (6)

Foremost in this field of contingency theories was, and is perhaps, Fred E Fiedler. Small in stature with a visage not unlike an irish terrier - and with a personality to match - this giant intellect from the University of Washington reduced the chaos of conflicting leadership studies to a semblance of order. He explained the apparent inconsistencies in his results in terms of moderator variables (7). First variable you recall was, 'the favourableness of the situation to the leader' (position power, task structure, relationships), Second variable was, 'the orientation of the leader' (task orientated or relationship orientated). This leadership style was to be measured by an ASO score (the assumed similarity of the opposites - least preferred co-worker and most preferred co-worker or LPC and MPC).

But you know and I know that, notwithstanding that explanation and admitting the importance of situational factors, there are some men who more often than most fill the leadership position in almost any group and almost regardless of place and time, whilst there are others who never do, nor ever will. There is it seems therefore a need to resist the over-emphasis of situational determinancy and still a need to identify the permanent sources of influence and prominence operating in a wide range of situations. So we go full circle, back as it were to the qualities approach. Without actually denying that which Fiedler has so persuasively written - who dares - there is a need to explain why some men are more frequently leaders, regardless of circumstance, whilst others are not. Let us then look, briefly, at this twin question of ability and of style.

But first, perhaps you will permit me to indulge in some intuitive experiential psychology, and be dogmatic. Purry (1973) says:

> "Leaders are all bright. Sometimes they are brighter than others, sometimes they are positively clever, sometimes shrewd, may be understanding and perceptive, possibly may be possessed of insight, can see wood for trees, can get to the roots of a problem, can devise alternative solutions to problems, occasionally may even be imaginative, but always, in one way or another, are, mentally bright"

---

6. Introduction to Thought and Personality, Ed Warr P 3, Penguin Books 1970.
7. See A theory of leadership effectiveness, Fiedler F E, McGraw-Hill 1967. For a briefer presentation see Leadership - a new model, Fiedler F E, Discovery 1965. Or see The contingency model and the effect of leadership training and experience, Fiedler F E, in Proceedings of NATO leadership symposium, Aug 1971.

In 1904, as we have seen, intelligence had but four elements, but the science of the subject, for that it quickly became, grew up (8). By the 1950's it was possible for Burt to say without fear of contradiction:

> "The degree of intelligence with which any child is endowed is one of the most important factors determining his general efficiency all throughout life" (9)

and for Vernon:

> "This means that, fairly good predictions of ability in education, industry or everyday life, can be achieved by 'g' tests alone, and that somewhat more ground can be covered by tests of the main group factors" (10).

And it is perhaps notwithout interest that, after the Central Committee of the Communist Party of the Soviet Union had addressed itself in 1936 to the merits and, more specifically, the demerits of psychometrics, psychological tests were banned de jure in the USSR, but they are now used de facto in several contexts including space research whilst among the tests are elements from the best seller of all time, Raven's Matrices. (11)

In short, we may say that intelligence, whatever it is, does, broadly speaking, predict potential occupational status.

But there is also more recent, and equally compelling, evidence that above certain cut-offs and within certain bands of intelligence, IQ does not predict. Guilford noted:

> "With about fifty intellectual factors already known, we may say that there are at least fifty ways of being intelligent" (12). (And many more of being stupid).

Not surprisingly then, though we may say intelligence predicts, with as much verismillitude, it may be said that, in some contexts, it does not. In particular, intelligence neither predicts excellence, nor, as we know from many military studies, is it of much help in the selection of potential leaders.

---

8. And not least, to be fair to Spearman, because of his pioneering work with tetrad differences and factors.

9. The evidence for the concept of intelligence, Burt C, British Journal of Educational Psychology, Vol 25, 1955.

10. The structure of human abilities, Vernon P E, 1950.

11. See, To test or not to test: trends in the Soviet news, Joseph Brozek, Journal of the History of the Behavioural Sciences, Vol 8, No 3, 1972.

12. Three faces of intellect again.

The paradox is resolved like this. As traditionally measured, intelligence in all its many and various forms was, despite that variegation, always measured as a 'convergent' ability. The tests had a correct solution which the testee had to find thus a premium was placed on conventional thinking and on conformity. In the fifties, however, there developed (and grew apace 'till it became a runaway bandwaggon) a whole new field of intelligence testing concerned with abilities of a different kind. Variously named, 'creativity', 'originality', 'divergent ability', 'lateral thinking' and 'cognitive complexity' these new found abilities or styles of mental process were thought to explain the variance not accounted for by the conventional (convergent) tests. Certainly there is a growing body of evidence which suggests this is so. (13)

This literature (or movement) had an important corollary in that it became inextricably mixed with parallel developments in studies of flexibility (versus rigidity), concreteness, ambiguity intolerance, dogmatism, conformity and authoritarianism - all it might reasonably be thought primarily matters of personality rather than of ability. So in this growing and indeterminate field, part ability, part personality, we find mixed together tests of risk taking, tests of category width, tests of remote associations, uses of an object, measures of fascist tendencies, Roe's eminent scientists, and the contrary imaginations of Liam Hudson's schoolboys whose highest scorer, you recall, wanted to stick nails through the sides of his barrel, put aunty inside, and roll her down a hill! Herein (not in the barrel) lies the key to the problem.

Logically, I suggest, ability and personality are not and cannot be, independent. The evil intent to put aunty into a position of discomfort must needs be matched by the inventiveness necessary to conceive the vehicle in which to do it. Similarly, a personality properly described as, 'mercenary, calculating, sharp and shrewd' cannot exist without the ability to make evaluations, manipulate monetary data, comprehend and assess relative advantage. The style is necessarily dependent on the capability. In the other direction, since, in general, cognitive skill or ability does not persist long in a vacuum of disuse, the ability to perform in a certain way depends on the practice implied by an habitual pattern of that behaviour which is generally called a 'disposition' or may be called 'personality'.

---

13. Names to look for in this literature are: Getzels and Jackson, Anne Roe, Mckinnon, Liam Hudson and on cognitive complexity, see Changes in students' performance on personality tests during university courses, Abercrombie M L J, et alia SSRC report, Annex 1 by Stringer P.

What I want to suggest is this. When Fiedler talks of leadership style he must necessarily imply the ability to behave in that way. Furthermore the ability which his ASO score measures (or at least taps) is a facet of cognitive complexity. Those leaders who Fiedler describes as task orientated, because in their perceptions of people they see wider differences (low LPC scores) than do those who are relationship orientated (high LPC scores), must have the ability to differentiate implied in the concept of cognitive complexity in order to be able to adopt that style. They make those discriminations because they are able to make them and have a choice whether they do so or not. Those without this perceptual acumen cannot behave in that way.

Moreover, I think it may be argued that a whole host of leader and manager functions depend on this critical ability to discriminate. Shrewd delegation of responsibility, and the perceptive and differential handling of men are but two examples dependent on this ability to make accurate discriminations. Therefore I think it possible that in this new arena of intelligence testing there is the potential to explain some elements of leadership by the use of an objective measure of this cognitive complexity. Let me describe how I tried to go about it.

Discrimination Test Construction

Officers newly, and for the first time, posted into positions of responsibility for training other, more junior, officers attend a course some of which is concerned with problems of appraisal. Apart from lectures, demonstrations, and discussions they are required as a training exercise in appraisal to write OERs on LPCs and MPCs. The data derived from these exercises are additionally used in a continuing programme of research into person perception and the factors affecting adequate performance appraisal. In particular, a difference score (the inverse as it were of Fiedler's ASO score) is calculated from this OER writing exercise by simply taking the sum (of the modulii) of the differences between LPC and MPC rating on each of the 27 rating scales in the OER. (14) These sums over 27 scales varied from as low as 31 to

---

14. For an erudite discussion of difference scores see Proposals leading to analytic treatment of social perception scores, Cronbach in Person perception and inter-personal behaviour, Ed Tagiuri & Petrullo 1958.

as high as 136.

After one week together the course members engage in some simple socio-metrics. Inter alia, a hypothetical project is described and a leader for it chosen.

Interestingly, correlations between these difference scores and peer selection as leader of the project have, with one exception, all been positive and significant. Relating this then to our thoughts on cognitive complexity, it seems at least possible that ASO scores or difference scores might be used as a simple and objective measure of leadership potential.

If the 'ability' and willingness to perceive differences in others is an attribute possessed by those likely to be selected as leaders in informal groups of officers by definition themselves experienced, competent and at least moderately successful (otherwise they would not have been selected to engage in officer training), then here is something - call it ability, style, personality, cognitive structure, perceptual acuity, what you will, pragmatically it matters not - which may be easily and objectively measured and as easily scored. But were simple LPC and MPC and the derived difference score to be used in the selection situation, there would be two obvious drawbacks:

Once the word was 'out', as very soon it would be, scores could and doubtless would be faked. All LPCs would rate; 1, 1, 1, 1, ... whilst MPCs would be halo 9 men.

Even if this should not happen, the test would lack standardisation. My LPC is not yours. Neither do we share an MPC. It could legitimately be argued that true differences between L and MPCs must necessarily vary.

It seems from this that a requirement may be stipulated as follows. The ability to differentiate between people may be measured by:

A standardised test which is;

Meaningful to the target population and which;

Exercises the 'ability' to perceive true differences and discriminate between people in the way that LPC/MPC ratings 'test' that ability. Furthermore it should;

Include 'cheat' detectors, ie scales for which no evidence is available and therefore IK is the correct assessment, and scales in which the relative positions of the persons being judged are reversed as compared with their positions in the majority of cases. And finally;

The test must have a known 'solution' derived from research findings of the relationships between behaviour and attitudes described in the test with subsequent performance either in training or in an employment environment.

And it must satisfy normal test criteria like stability and reproducibility and so on.

Since, in the long run, use of this particular test was intended for a particular selection centre, the intention was formed to construct a narrative type description of two candidates at that centre, ie persons in exactly the same position as the target population, in which information could be used by the perspicacious subject to discriminate between them. The information in the test was to be obtained by reference to research findings so that there would be a DS answer.

Such research evidence exists in American literature in abundance. You will recall the famous question, "Did you ever build a model aircraft that flew?", which is reputed to have been almost as good a predictor of success in flight training during World War 2 as the entire air force battery. But how far one is entitled to take such evidence across the cultural divide between one English speaking nation and another, is a matter for conjecture and possibly doubt - though recent cross-cultural studies have suggested considerable similarities in the value systems of England and America (15). Necessity being the mother of invention, it will be shortly shown that compromise was made. English evidence actual derived from the selection situation in which the test might be used was relied on to the extent of its availability whilst findings from other environments provided supplementary material.

Smith and Jones

Perhaps before going on to describe the substance of the test, I may briefly remark on the apparently simple matter of putting a name on it. "What's in a name? That which we call a rose by any other name would smell as sweet", is poor counsel to the test designer engaged in finding a title for his instrument. He may wish positively to mislead his subjects - as is so frequently done when personality measures are thinly disguised as performance or ability checks - or (and in my view more properly) he may merely wish to avoid gratuitous antagonism at the commencement of his work and, by the selection of a neutral title, happily and hopefully would wish to minimise bias, prejudice, and/or emotional set. In any case, some care is necessary.

---

15. See Personal value systems of managers - so what, England G W, a paper to the Seminar on cross-cultural studies, Brussels 1973. And Accelerated manager's objectives in twelve countries, Bass BM et alia in Industrial Relations, Vol 12, No 1, May 1973.

In England, if not elsewhere, the names 'Smith' and 'Jones', especially in simple juxtaposition just like that, are almost synonymous with, 'Mr Average', 'the man in the street', or as English law has it, 'the man on the Clapham omnibus'. The names are emotionally neutral and flat: certainly they are equal - rather like the advocates', Roe and Doe. By contrast, to have chosen, 'Bannister Mainwaring-Smythe' for appraisal alongside simple 'Jack Jones' might well invoke the intrusion of social class overtones and invite bias - error variance without control.

So simply, 'Smith and Jones - Judgement Exercise' it is and honestly described. (16) Instructions are given as follows:

> "Smith and Jones are two officer candidates. Both have been tested, interviewed, and extensively observed in small-group and individual exercises. Your job is to make an assessment of each using the scales provided and finally to decide whether either or both should be accepted for officer training....."

Instruction on the use of rating scales is given in a manner not unlike that given in a number of other studies in this field in which two persons are rated using one form. The rating scales to be used are given and include such dimensions as: motivation, academic potential, sociability, and so on.

The narrative then describes Smith and Jones simply and sequentially as the information would be obtained and recorded in the selection procedure with which my experimental officers are familiar - having been through it - and with which the potential users of the test would be becoming familiar because the test is intended to be a part of it. The narrative has five major parts of which I will describe just three:

Information obtained at first interview
Results of paper and pencil questionnaires
Observation of the leaderless group discussion exercise.

---

16. For interesting discussions of the very serious issue of privacy and research see SSRC Newsletter 18 Mar 73, the Younger Report Cmnd 5012, London HMSO 1972, and generally on the ethics of selection see The Psychology of Selection, Wilson J H in Management Today, Jan 1973.

Interview Data

From the first interview it is learned that:

"Aged 19 1/12, Smith is the fourth and youngest soon of a self-employed electrician.....

"At his local comprehensive school he obtained (examination passes) in ..... at age .....

"At (age) he passed (higher level examinations) in .....

"He played (various games at team levels) .....

"Was a school prefect .....

"And an active member of a local youth club ....."

And then Jones' background is similarly described.

We know from the McGill studies (17) of the weight attached by PSOs to achievement data, whilst, more recently Martin (18) at CAF PARU has confirmed this interest:

"Officer applicants tended to be assessed on the basis of academic achievement about which interviewers, it seems, could agree"

"Academic achievement correlates highly with military potential ..... the PSO's prediction of a candidate's potential for remaining in the Service and progressing in rank is based largely on the candidates academic achievement at the time of the interview".

Moreover, the evidence reviewed by Asher (19) of the better-than-most predictive value of historical and verifiable biographical data suggests this strategy both prudent and wise.

---

17. Decision making in the employment interview, Webster E C, Industrial relations centre, McGill University 1964.

18. Reliability and validity of CAF Selection interview, procedures, Martin H A, Report 72-4, DFPARU, May 1972.

19. The biographical item: can it be improved? Asher J J, Personnel Psychology 1972, 25.

It therefore seemed reasonable and fair that the opening paragraphs in "Smith and Jones" should provide a simple resume of family background, scholastic achievement, membership of and status in clubs, societies and para military pre-service units. It is also true of course that the statistical association of these variables with success in officer training in my own organisation is known. Thus there is a OS 'solution' to this part of the problem.

The narrative goes on to build a picture of the preferences and expressed values of Smith and Jones.

"Questioned closely about his attitudes and interests Smith said that he liked .....

"He gave the impression that ..... were of importance to him .....

"He seemed to prefer .....

Jones:

"Jones liked .....

"Although he had done neither he thought he would like .....

"He expressed a distaste for ..... and so on.

Research using Gaygier's likes and interests test had previously established significant correlations between the expression of these preferences by candidates and subsequent success in training. Thus there is again a 'correct' solution.

Questionnaire

Similarly the association between success in training and the answers given to an attitude questionnaire specially designed to measure 'motivation' is known. (40) The Smith and Jones narrative therefore includes a simple tabular presentation of some of these questions and the answers supposedly given by Smith and Jones.

Leaderless Group Problem Solving Discussion Exercise

The narrative also includes a summary of the way in which Smith and Jones are supposed to have behaved in a leaderless group problem solving discussion exercise. Doubtless you will be familiar with some of this material:

---

40. For a detailed account of this and other personality measures see The Measurement of Motivation in the Selection of Aircrew, Parry J B, Science 4 (RAF) Memo 157, December 1968.

"..... spoke frequently during the discussion (21) .....
"his contributions were largely procedural suggestions (22), eg 'Let's have each person give his solution', seeking information, eg 'Does anyone know .....', and integrative, eg 'We seem to agree that .....', and 'This would modify what No 3 said earlier' " (23)

All this information so far discussed goes the same way. That is to say, the 'goodie' has the better achievement data, (although it is not too easy to see this and in discussion there is considerable argument about its weight), whilst his likes and dislikes, attitudes and motives are those associated - as far as we know - with 'better' candidates. Further, his supposed performance in the small group discussion and in the hangar exercises (which I have not described) is intended to be compatible with prominence in these situations and with the exercise of influence. But lest this make the discrimination task unduly easy the matter is deliberately made ambiguous by the injection very early in the narrative of descriptions of physical and facial appearance known to have a diametrically opposite effect to that of the personality and performance material.

In 1955 Johnson (24) did some interesting research in which she described to subjects two opposites. The 'baddy' was ruthless, brutal, hostile, boorish, vulgar and so on, whilst the 'goodie' was warm-hearted, honest, intelligent, responsible, etc. She then asked her subjects to rate the likelihood that these two would have certain physiognomic features and found agreement between her S's that her goodie and baddy would indeed look significantly different. To inject some potential ambiguity into Smith and Jones I therefore attributed Johnson's baddy to my goodie and vice versa. (And it is extraordinary how many of my subjects are very much confused by this purely physiological and physiognomic data).

---

21. See An Analysis of the leaderless group discussion, Bass B M, Journal of applied psychology, Vol 33, on participation time and influence.

22. On procedural suggestions, see Some Factors in the selection of leaders by members of small groups, Kirscht et alia Journal of abnormal and social psychology, Vol 58, 1959.

23. On integrative behaviour and leadership see The Behaviour of Leaders and other group members, Carter L et alia, Journal of abnormal and social psychology, Vol 46, 1951.

24. See Stereotypes and perceptual distortion, an unpublished honours thesis, Mary Washington College 1955, Johnson L, extensively quoted by Secord in Person Perception and Inter-Personal Behaviour, Ed Tagiuri and Petrullo, 1958.

Discussion

Just 8 courses, approximately 150 officers, have now attempted this exercise. A scoring schedule permits minor discrepancies from the correct solution but penalises the following:

Smith and Jones in reversed positions on any scale
Either Smith or Jones the wrong side of the half way point on any scale
Either Smith or Jones more than 1 rating in error in either direction
Either Smith or Jones given gradings on scale for which there is insufficient information to make a judgement.

Results so far may be summarised as follows. Just under 1/5 of subjects have totally misread the situation and preferred the 'wrong' candidate. The majority prefer the 'right' candidate but the accuracy with which this preference is expressed in the ratings varies. There is a positive correlation between these Smith and Jones scores and the difference scores obtained in the LPC/MPC exercise. There is a positive correlation between the Smith and Jones scores and selection by peers as leader of the hypothetical project.

It is therefore concluded that there is a suggestion, and I put it no higher than that at this stage, that there is a trait/ability to perceive differences between people and to use scales to record those perceptions. This ability to differentiate may be causally related to leadership potential. At least they appear to co-exist and be statistically correlated. Tests designed on the principles which I have outlined may then have a potential as objective measures of this perceptual ability which seems to be one facet of the cognitive complexity one would expect to find in a leader.

It is early days yet. These first gropings after a little understanding cannot and do not warrant more than the most tentative of conclusions, but of one thing I am sanguine: as a training exercise, if not as a test, Smith and Jones has considerable potential. Subjects who have worked with it have testified to its realism. A common response is, "I had Smith in my squadron", and "presumably these are not their real names". A small minority in every group so far making a totally different judgement from those of their colleagues, the fact of assessment unreliability is forcibly demonstrated to participants. One of the problems of appraisal is thereby meaningfully brought home. (And it is my experience that this recognition that a problem may exist is a necessary precondition for training in this difficult matter). The material provides an excellent common experience for reasoned discussion on inter-person perception and judgement whilst many of the lessons which must be learned if appraisal

is to improve may be drawn from it. Recalling our Washington meeting: Wilkinson (25), reporting a USAF review of evaluation systems emphasised the need to develop methods for training raters. I commend 'Smith and Jones' to you as one such method.

In conclusion then, I have tried to explain how I think one of the aspects of the 'new' intelligence may perhaps be related to that potential for leadership which exists more in some men than in others. I do not claim to have developed yet a usable test of this trait/ability but I do suggest that the line of enquiry and development I have outlined is worth following in that context. I commend a Smith and Jones type exercise to you if indeed you should feel that the problem of appraisal in its widest sense is more to be solved by training than by the (oft repeated) construction of new forms so that we may enhance the utilisation of human resources in conditions of growing demands. And finally, I would welcome your guidance and comment both formally and informally; and thank you for your patience in listening.

---

25. USAF evaluation systems: review and research recommendations, Wilkinson R E, Paper presented to MTA Sept 1971 in Proceedings of the 13th Annual Conference published by the US Marine Corps, Washington DC.

New Concepts for the Measurement of Attitudes and Motives[1]

Robert G. Smith, Jr.
Human Resources Research Organization

INTRODUCTION

Attitudes and motives seem to represent especially difficult problem areas for applied psychologists. We can measure opinions and beliefs with great precision. Yet our subjects often do things they say they won't do, or dislike to do, and vice versa. And even when it is perfectly clear that they "oughta wanta" do something, they don't.

Frankly, this is a theoretical paper. It is theoretical, not from the sense of a precise formulation, but from the standpoint of a way of looking at problems. Some ways of looking at problems are more fruitful in yielding potential solutions than others. I hope to present some ways of looking at attitudes and motives that will yield more helpful solutions to difficult problems than others.

The thesis of this paper is that these problems stem from our ways of thinking about attitudes and motives -- defining and conceiving of them. Once we have appropriate ideas about attitudes and motives, then we can both measure and change them.

---

[1]Paper to be presented at the meeting of the Military Testing Association, October 1973.

TWO WAYS OF LOOKING AT ATTITUDES

It is difficult to discuss attitudes without getting controversial, or saying something that sensitive people object to. So, bear with me while I weave a fantasy.

Let us imagine that the U.S. program for space exploration has established contact with the third planet of another star. This planet is populated by people who are very similar to us on earth, except for one characteristic. They are purple. After several centuries of trade, travel, there are scattered throughout earth small groups of these purple people, known as the Pankees.

On the whole, relations between earth people and the Pankees are neither better nor worse than relations between different groups of earth people -- sometimes so-so, sometimes terrible. There is, however, one group which in its hate for the Pankees is notorious. Because this group identifies itself with a specially designed handkerchief, it is called the Hankee Pankee League (HPL). Members of this league do the following with great frequency (sometimes described as a purple passion).

a. Use highly insulting language when referring to Pankees.

b. Write Pankee Go Home on walls.

c. Avoid contact with Pankees.

d. Burn Purple crosses on Pankee front lawns.

e. Beat individual Pankees, thus spawning that ancient folk song about the Purple People Beaters.

f. Go into a fearful rage whenever Pankees give their salute, a sign of solidarity. (The salute is a raised clenched fist, with the middle finger extended.)

Earth psychologists, after careful study of the situation, have made pronouncements about members of the HPL. *Some psychologists* say they possess negative attitudes toward Pankees. Still others say they exhibit a wide range of Anti-Pankee behavior. Which one of these pronouncements you adopt will make a great deal of difference. Let us see why.

(Insert Figure 1 About Here)

Figure 1 diagrams this situation. The organism -- according to our fantasy, an HPL member -- receives stimulation related to Pankees. He responds by insults, writing on walls, etc. As long as we refer to these responses collectively as Anti-Pankee Behavior, we should have relatively little difficulty in dealing with them. If we use the terms Prejudice and Attitude as simple ways of referring to the total collection of responses, we are still in good shape.

(Insert Figure 2 About Here)

However, we are very likely to do the following, as shown in Figure 2. We are very prone, conceptually, to put the attitude or *prejudice inside the* individual, and say he <u>has</u> an attitude or prejudice. Once we do this, we have created all kinds of *problems for ourselves*. Since the only benign way of approaching entities within the individual is with words, we begin to try to measure attitudes and prejudices with words.

The history of verbal attitude measurement certainly proves that if you ask the same or similar questions over and over again, you can get highly consistent responses. In fact, most of the more common methods of attitude measurement are based on this fact (Edwards, 1957).

It is, unfortunately, also well known that verbal behavior, no matter how consistent, frequently is not consistent with other behaviors. Despite this, the area of attitude measurement is still devoted in large measure to the development of verbal consistencies.

TWO WAYS OF LOOKING AT MOTIVATION

As long as I have been associated with trainers and educators, I have heard them plead for more highly motivated students. What do they mean by this? What is a motivated student?

As you talk to training personnel, you find that they describe this student in the following general terms:

A motivated student --

Stays awake and alert during lectures.

Works steadily during laboratory or practical exercise periods.

Reads additional books on the subject.

Tells the instructor how interesting the subject is.

Voluntarily spends free time in the instructional area.

In other words (primarily those of Mager [1968]), the student tends to approach the subject matter rather than avoid it.

(Insert Figure 3 About Here)

Figure 3 diagrams this situation. The student, stimulated by the school situation, responds by a variety of forms of approach behavior. If we use the term motivation as a shorthand term to describe all those behaviors, we can deal with the situation in an effective fashion.

(Insert Figure 4 About Here)

On the other hand, if we do as shown in Figure 4, and conceive motivation as being inside the individual, we have effectively removed it from our view, and are limited, as in the case of attitudes and prejudice, to verbal approaches.

THE IMPORTANCE OF BEHAVIOR DESCRIPTION

I am convinced that a good case can be made for the thesis that the key to major advancement in almost any field of applied psychology is the description of behavior and the cues that elicit or affect it.

The techniques of behavior description for man-machine systems pioneered by Miller and others (1962), opened the door to major advances in the precision with which manpower planning for these systems can be conducted.

Computerized methods of analyzing data concerning job duties and tasks, pioneered by Christal, Morsh (1964) and colleagues have opened the way to major advances in personnel management.

The most important discovery that we in HumRRO made concerning the development of effective and efficient training was that what is taught (the behavior required of the student after instruction) is generally more important than how it is taught. This discovery had profound consequences in the development of programmed instruction and the systems approach to training and education, all of which emphasize the careful description of the behavior required of the student (Smith, 1971).

The people comprising the behavior modifiers, applying Skinnerian concepts to a wide range of problems have developed many new ways of dealing with the maladjusted, autistic children, school children, and many others. Again, the key has been the careful description of the behavior involved.

It is because of this history that I feel it is important to describe the behavior associated with the concepts of attitude, prejudice, and motivation. This behavior can then be the foundation for new and more effective ways of measurement.

TOWARD BEHAVIORAL DESCRIPTION IN THE AREA OF PREJUDICE

As a beginning for the description of behavior related to prejudice, let us turn to Allport's (1954) analysis. Allport identifies five general classes, or degrees, which prejudiced behavior may take. These are Antilocution, Avoidance, Discrimination, Physical Attack, and Extermination. For ease in communication I will call the object of prejudice the other.

Antilocution refers to talking against the other. With persons who think the same way he does, or on anonymous questionnaires, or sometimes with strangers, the prejudiced person freely expresses his antagonism, opinions and beliefs.

Avoidance. A more intense form of behavior is to avoid members of the other group. Even at the expense of inconvenience, the individual removes himself from situations in which he will encounter the others. He does not directly harm the other groups, but simply withdraws.

Discrimination. This is more active than avoidance. Now the individual actively makes distinctions that are harmful to the other. He may try to keep others from certain jobs, residential areas, hospitals, schools, churches, or various social privileges.

Physical Attack. Under conditions of high fear or anger, acts of actual or symbolic violence may be carried out against others. Crosses may be burned; others may be beaten or forced to leave the forbidden places; insults may be painted on walls; the others' places of worship may be bombed.

Extermination. The final and tragic level of prejudice is the killing, lynching, or extermination of the other.

Allport believes that these forms of behavior represent a rough scale -- that activity on one level makes it easier for prejudiced behavior to move to a higher level. This is one reason why it is important to measure the behaviors involved at the various levels.

What can we do to measure the various behaviors involved in prejudice?

For antilocution, we can take an up-to-date inventory of opinions and beliefs concerning the other. These can be obtained by interviewing both the individual and the other. It would try to express these critical comments as clearly as possible, but I would not subject them to any attitude scaling technique. Instead, I would ask respondents if they had heard any of these comments expressed within a short time period. In this way, a general index of the amount of anti-locution going on could be obtained. Periodic administrations could then alert commanders to the rise or fall of anti-locution within their commands, and give clues as to the effectiveness of race relations programs or the possibility that prejudicial behaviors are beginning to escalate to the next highest level.

For avoidance, I would identify the most common situations in which avoidance might occur. I would then use this inventory to identify the general levels of both avoidance and positive social contact. My suspicion would be that the others would be more sensitive to avoidance than the individual himself.

For discrimination, it is recognized that many of the most overt and obvious forms have been outlawed. Yet, from all that I have seen and read, there are still a number of more subtle discriminations remaining. Again, I would develop an inventory of common forms of discrimination, and I would rely more heavily upon information from the others, since they would recognize many forms that the individual would not.

For physical attack, again, I would develop an inventory of the more subtle and hidden forms of property destruction, insults, and

assault. Those incidents that attract the attention of the commander or the military police, and often the press, are easily recorded, but are often the "top of the iceberg."

Extermination is clearly not a national policy, as it was against the Jews in Nazi Germany, or against the Hutus in Tutsi-controlled Burundi. Yet, the race riots and mass fights that have occurred on military vessels and posts border on attempts to exterminate small groups.

Now, let me sum up. I believe that prejudiced behavior can be measured through the inventory approach -- not through the attitude scaling approach -- and I urge the services to develop and apply such inventories on a regular basis.

There is a model for what I have proposed. That is the HumRRO effort to develop a list of discriminatory and irritating behaviors which U.S. troops and commanders displayed (often unknowingly) toward Katusas. These were Korean soldiers serving in U.S. units (McCrary, 1969). This work suggests the possibility of using the inventories as a means of informing commanders and others exactly what their responsibilities are in the area of control of prejudiced behavior.

In fact, I am convinced that the approach to measurement that I have described holds the key to the development of practical methods for the elimination of prejudice. But that is another paper.

TOWARD BEHAVIORAL DESCRIPTION OF STUDENT MOTIVATION

Mager (1968) has suggested an approach to the behavioral definition of attitude toward learning. A student has a favorable

attitude, or is positively motivated toward learning if he approaches the subject or the learning situation. Conversely, he has an unfavorable attitude, or is negatively motivated toward learning if he avoids the subject or the learning situation.

For suitable measurement of the motivation of students, we need to develop an inventory of approach behaviors and avoidance behaviors common to students in a particular course. To be most useful, the inventory should be specific. Then the inventory could be used by either instructors or students.

Once these approach and avoidance behaviors have been identified, then we can change the student's environment so that approach behaviors are reinforced and avoidance behaviors are extinguished. Mager presents a number of ideas for accomplishing this.

SUMMARY

To summarize my main point. We can make good progress toward dealing with problems of attitudes and motives if we conceive of them as shorthand descriptors of appropriate behavior. We will make little progress if we conceptually bury them inside the individual, therefore making them inaccessible except through the medium of words.

Figure 1

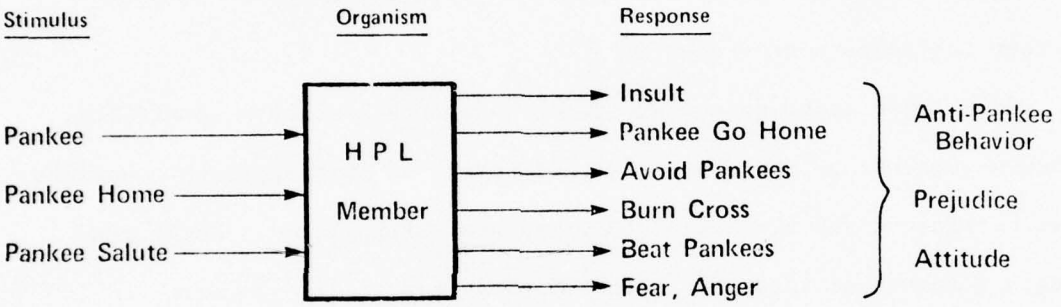

Stimulus
Organism
Response
Pankee
Pankee Home
Pankee Salute
H P L
Member
Insult
Pankee Go Home
Avoid Pankees
Burn Cross
Beat Pankees
Fear, Anger
Anti-Pankee Behavior
Prejudice
Attitude

Figure 2

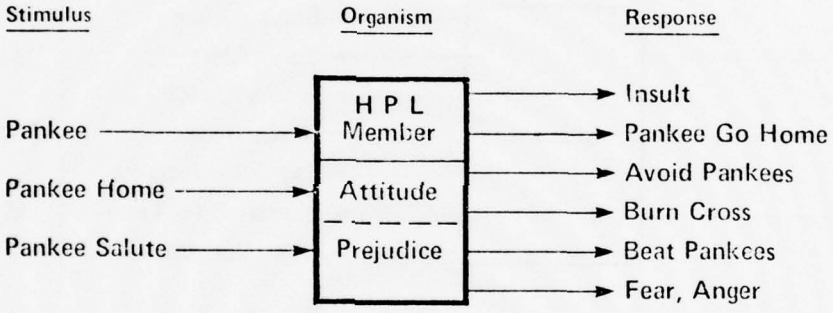

Stimulus
Organism
Response
Pankee
Pankee Home
Pankee Salute
H P L
Member
Attitude
Prejudice
Insult
Pankee Go Home
Avoid Pankees
Burn Cross
Beat Pankees
Fear, Anger

Figure 3

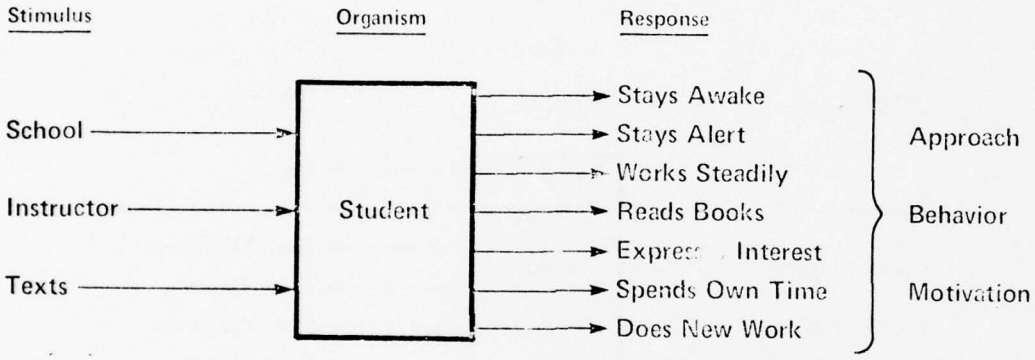
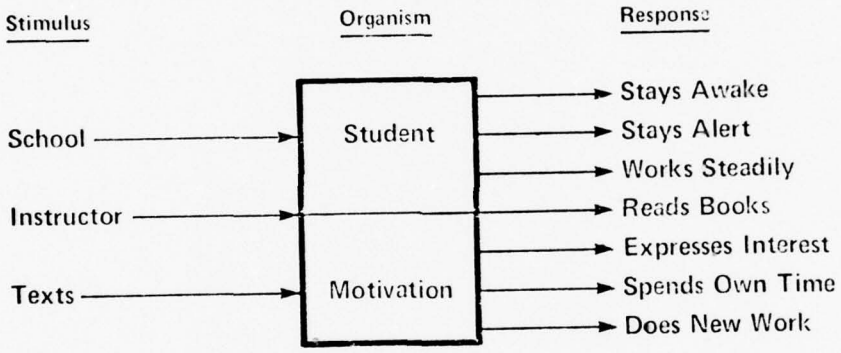

Stimulus
Organism
Response
School
Instructor
Texts
Student
Stays Awake
Stays Alert
Works Steadily
Reads Books
Expres Interest
Spends Own Time
Does New Work
Approach
Behavior
Motivation

Figure 4


Stimulus
Organism
Response
School
Instructor
Texts
Student
Motivation
Stays Awake
Stays Alert
Works Steadily
Reads Books
Expresses Interest
Spends Own Time
Does New Work

REFERENCES

Allport, Gordon W. The Nature of Prejudice. New York: Addison-Wesley, 1954.

Edwards, Allen L. Techniques of Attitude Scale Construction. New York: Appleton-Century-Crofts, 1957.

Mager, Robert F. Developing Attitude Toward Learning. Palo Alto, California: Fearon Publishers, 1968.

McCrary, John W. Jr. Commander's KATUSA Program Checklist, Alexandria, Virginia: HumRRO, 1969.

Miller, Robert B. Analysis and Specification of Behavior for Training, in Glaser, Robert (Ed.) Training Research and Education, Pittsburgh, Pennsylvania: University of Pittsburgh, 1962.

Morsh, Joseph E. Job Analysis in the Air Force. Personnel Psychology, 1964, 17(1), 7-17.

Smith, Robert G., Jr. The Engineering of Educational and Training Systems. Lexington, Massachusetts: D.C. Heath, 1971.

ADMINISTRATION OF MULTIPLE-CHOICE TESTS to NON-READERS VIA TAPE RECORDER: Case Studies

Joseph L. Boyd, Jr., Ed.D.
Educational Testing Service

The Automobile Mechanic Certification Tests are administered by Educational Testing Service for the National Institute for Automotive Service Excellence. Tests are given twice a year at some 200 test centers in all fifty states. The Program offers voluntary certification, based on a combination of work experience and test performance, in various service areas, i.e., "Engine Repair," "Brakes," "Heating and Air Conditioning," and as "General Automobile Mechanic" for those passing the tests in all areas.

However, my objective today is not to brag about the program, but to discuss a problem that faces us and to tell you some exploratory steps we are taking to solve it. The problem is literacy--the ability to read the test questions. Our initial approach was that no individual could be a competent mechanic if he was unable to read the automobile manufacturers service manuals. One principle of test development we adopted was that the reading and vocabulary of the tests should be no more difficult that the "well-thumbed" sections of the service manuals. Not so. We soon discovered there are two groups of mechanics, many highly esteemed by their employers, who could not be adequately tested by conventional paper-and-pencil means. The first group consists of English-speaking mechanics who cannot read. Some are marginally literate, but they are placed at a great disadvantage when attempting to read a test. The second group consists of mechanics whose native language is not English. They may be literate in their own language but not in English.

The traditional method of accommodating this type of test-takers is to allow a translator or reader. We rejected this approach because we were concerned with test security problems that might arise, and because a satisfactory translator or reader would have to have a

knowledge of the automotive technical vocabulary and could inject some of his knowledge, unintentionally perhaps, into the testing process.

In some situations, it is feasible to translate the tests into the foreign languages. For example, the Electronics Inventory Examination of the Guided Missile School at Redstone Arsenal was translated into German, Greek, Chinese, and Turkish. This was acceptable since these translated tests were to be used at only one location and with sizeable groups. On the other hand, for the automobile mechanic tests, the calls for assistance came from every part of the country, and for an unpredictable mixture of language--Greek, Spanish, French, Italian, and would you believe, Finnish!

However, the greater need seemed to be to help the non- or poor reader. The solution we arrived at was to record the text of the test so that each mechanic could listen to the words while following along in his test book. Each candidate would have a cassette player and a light weight headset. He could reverse the tape to hear the question read as many times as he needed. Most of these tests allow three hours of testing time but the test text fitted easily on a sixty minute cassette. The one test which was timed for four hours required a second tape. The handout shows five typical test questions. This is what a mechanic would hear. (PLAY TAPE OF FIVE TEST ITEMS.)

This system was tried out at the regularly schedule test administration on April 28 and May 5 of this y-ar. It was simply a try-out of the feasibility of the system--certainly not a controlled experiment. Four mechanics whose employers stated they could not read well enough to take the test participated. In addition, we had two other mechanics whose native language is not English who volunteered to take the tests. These men felt their grasp of spoken English was good enough to try the tests. Let us take a look at the results, case by case.

J. D. of Canton, Ohio took the tests in the fall of 1972, and failed all tests. His percentage scores were at, or barely above, the chance

level: 34, 38, 35, 24, and 26--for 4-choice multiple choice items. Although he did not pass any of the tests in the spring administration, his scores were substantially improved. There were seven scores. His percentages were: 61, 63, 33 (corresponding to his earlier 35) 53, 60, 48, and 36. The test administrator was enthusiastic. He reported that J. D. worked his way through the tests fairly rapidly, though he did repeat a number of items. He also observed that the candidate was apparently following the text in his testbook with his finger as he heard it read on tape.

A.K. was another mechanic who had previously taken the tests. In the Fall of 1972, he scored 34 percent on "Front End and Brakes" and 24 percent on "Electrical Systems." In the spring series, using the tape, he scored 60 percent on "Front End," 48 percent on "Brakes," and only 34 percent on "Electrical Systems"--a substantial improvement in two out of three areas. One might conclude that A.K. really does not know much about "Electrical Systems." A. K.'s native language is Burmese.

The four other mechanics who took the tests with the aid of cassette tapes had not taken the tests in the Fall 1972 administration. Two of these candidates attained high enough scores to be certified in a service area: J. L. as an "Automatic Transmission (Light Repair) Mechanic" and R. B. as a "Front End Mechanic." Three of this group of four men scored 50 percent or more on each test they took--two times the chance score. My conclusion was that (1) the candidates did know something about automobile repair work, and (2) the candidates were able to comprehend the tests reasonably well through the process of reading, listening, or a combination of both.

The final case was G. S. who requested to be tested in Italian. However, he agreed to take the test with the English language tape. The test administrator reported the G. S. found it "difficult to concentrate on questions being read while at the same time trying to follow wording in his test booklet and/or operating the tape recorder--the use of the tape recorder became so cumbersome that after the 18th question, we eliminated the recorder and read the remainder of the test to him."

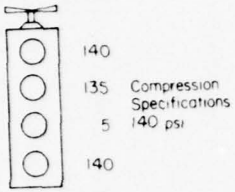
Five of the six men who took the tests in this way regarded the tape as a definite help in allowing them to demonstrate their competence as automobile mechanics. The Test Administrators concurred with the candidates' reactions.

We propose to use the tape-recorded test procedure again this fall, concentrating on mechanics who have previously taken the tests. Some time in the future, if the demand develops, we will record an oral translation of the tests into Spanish for those mechanics who are more comfortable in that language.

AUTOMOBILE MECHANIC CERTIFICATION

TESTS -- SAMPLE QUESTIONS

1.



After obtaining the compression readings shown above, the mechanic performed a wet compression test. The second set of readings was almost the same as the first. These tests indicate which of the following?

I. A burned value

II. A broken ring on a piston

(A) I only (B) II only (C) Either I or II (D) Neither I nor II

2. **The engine of a car with an automatic transmission starts in all selector positions. What is the most likely cause?**

(A) A shorted ignition switch

(B) A shorted neutral switch

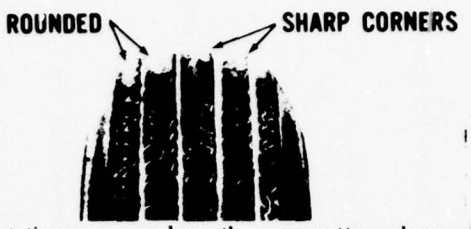
(C) An open neutral switch

(D) An open ignition switch

Copyright 1972

Educational Testing Service
Princeton, N.J. 08540

3.



Both front tires on a car have the wear pattern shown above. The sharp edges of the treads face toward the center of the car. This indicates that the front end has too much

(A) positive camber

(B) negative camber

(C) toe-in

(D) toe-out

4. ***If the accelerator-pump inlet check* ball is accidently left out of a carburetor, which of the following is the most likely result?**

(A) Flooding, especially when accelerating

(B) Hard starting and hesitation whenever the throttle is opened suddenly

(C) Hesitation or stalling whenever the throttle is closed suddenly

(D) Excessively rich mixtures for all low-speed, part-throttle, light-load conditions

5. **A mechanic is adding refrigerant to an air-conditioning system while the unit is operating. Where should he add the refrigerant?**

(A) To the high side only

(B) To the low side only

(C) To both the high and the low sides

(D) To either the high or the low side but not to both

Racial Differences in AFQT, AQE, and WAIS Scores of Air Force Basic Trainees

Stephen Knouse, David F. McGrevy, and Ronnie A. Thompson

Personnel Research Division, Air Force Human Resources Laboratory

Much recent emphasis in testing research has focused upon unbiased or "culture free" assessment. The problem of test bias against minorities is especially important to today's military which now must look to true volunteers for its entire manpower supply. In the past, draftees and those induced by the draft to volunteer produced a manpower pool with a sufficiently wide variance in aptitudes to provide adequately for personnel needs. Now, in the absence of draft pressure, the Air Force may be restricted in its manpower input and must insure its selection tests are tapping the full range of ability, particularly in minority groups.

The two primary selection tests employed by the Air Force are the Armed Forces Qualification Test (AFQT) and the Airman Qualifying Examination (AQE). Administered by the recruiter, the AQE yields four composite scores--Administrative (AQE-A), Electronic (AQE-E), General (AQE-G), and Mechanical (AQE-M). These Aptitude Indexes (AIs) are used in the recruiting and initial assignment of basic airmen.[1] The AFQT is primarily used to determine mental category of

[1] Note. On 1 July 1973 the Armed Services Vocational Aptitude Battery (ASVAB) replaced the AQE in the Air Force testing inventory. Like the AQE, the ASVAB provides the Administrative, Electronic, General, and Mechanical Aptitude Indexes. Intercorrelations between the ASVAB and AQE range from .59-.62 for Administrative, .72-.75 for Electronic, .71-.73 for General, and .71-.74 for Mechanical, based upon four pre-selected samples of 1000 subjects each.

pre-inductees. The test evaluates verbal, numerical, spatial, and mechanical abilities, and yields a single score. Administration of the AFQT usually occurs at the Armed Forces Entrance and Examination Station.

Two recent studies illustrate the direction of Air Force research into minority group test performance. The problem of racial bias in the AQE was investigated by Shore and Marion (1972) who employed the AQE to predict performance on the Specialty Knowledge Test (SKT), a test of job knowledge used in determining an airman's promotability to a higher grade. Bias, defined as underprediction of the SKT by AQE AIs, was not found in the 16 career areas investigated.

Guinn, Tupes, and Alley (1970) reported a study exploring the relationship of demographic variables on Air Force basic trainees to several aptitude tests. They found a majority of test scores were significantly related to race, educational level and geographic area of the trainee. Racial differences were most prominent in tests requiring specific prior knowledge, such as verbal and mechanical tests, while situationally defined tests measuring non-verbal, spatial, and psychomotor abilities revealed less pronounced racial effects.

A further investigation exploring the relationship of the AQE and AFQT to a test of general mental ability appears necessary in light of the evidence presented by these two studies. The Wechsler Adult Intelligence Scale (WAIS), an individually administered test of general mental ability, offers the Verbal (VIQ) and Performance IQs (PIQ) as well as the Full Scale IQ (FSIQ), against which the four AQE AIs can be contrasted. In the present study, the WAIS was administered to a group of black

Air Force recruits and a group of white Air Force recruits who had taken the AFQT and AQE prior to their entrance on active duty. Differences between the WAIS, AQE and AFQT scores were explored as well as the predictive relationship utilizing the WAIS IQs as criteria and the AFQT or AQE AIs as predictors.

Method

Subjects: The subjects were 200 male Air Force basic trainees. One hundred black and one hundred white airmen at Lackland AFB were randomly selected over a 12 month period (Feb 1972- Feb 1973) during the routine testing occurring on the sixth day of basic training. The age of the subjects ranged from 17 years to 25 years with a mean of 19 years. Education ranged from 8 to 16 years of schooling completed. Mean education level was 12 years.

Procedure: The Wechsler Adult Intelligence Scale was administered to the trainees by the authors. Because the authors were members of the military and the trainees might perceive them as representatives of the authority structure and similar anxiety-producing stimuli, military aspects of the situation were de-emphasized as much as possible. Each subject was asked to relax, to smoke or drink a Coke if he wished, and informed he need not call the test administrator "Sir." Each subject was further instructed that he was in an experimental situation and that his test score would not be entered into his records or used for job placement in the Air Force. Each WAIS record form was scored by the test administrator and checked by the other two authors.

Scores on the Armed Forces Qualification Test and Airman Qualifying Examination were obtained from basic training flight records.

Results

Means and standard deviations of the AFQT, the four AQE AIs, the three WAIS IQs, and the eleven WAIS subtests for the black and white samples appear in table 1.

Table 1 about here

---

t tests (df=198) revealed that blacks scored significantly lower than whites on the AFQT, AQE AIs, and the VIQ, PIQ, and FSIQ of the WAIS. Analysis of the WAIS subtests demonstrated blacks scoring significantly lower on all subtests with the exception of Digit Span and Digit Symbol, where there were no significant differences between blacks and whites.

A correlation matrix of AFQT, AQE, and WAIS scores appears in table 2 for the black sample and table 3 for the white sample. All correlations for the black

---

Tables 2 and 3 about here

---

sample were positive and significant ($p < .05$), except PIQ vs AQE-A, VIQ vs AQE-M, and FSIQ vs AQE-M. All correlations for the white sample were positive and significant ($p < .05$). The intercorrelations of the VIQ, PIQ, and FSIQ for the black sample were of a magnitude equivalent to that of the white sample. The intercorrelations of VIQ, PIQ, and FSIQ with the AFQT and AQE composites, however, were consistently much lower for black airmen than those of the white sample.

Regression equations were run utilizing the WAIS IQs as criterion scores and the AFQT or AQE AIs as predictors to determine racial differences in test scores. The three models employed are listed in table 4.

Table 4 about here

Table 5 represents a summary of the regression analyses.

Table 5 about here

Racial differences were found in all regression problems except the case of AFQT predicting VIQ. In the cases where PIQ and FSIQ were criteria for the AFQT and AQE-A and where PIQ was the criterion for AQE-G, the type of difference was a race main effect (no slope difference between black and white groups). All other differences were determined to be interactions between race and predictor test score.

Discussion

The results of this study indicated that only the AFQT in relation to the WAIS Verbal IQ demonstrated no racial differences. The AFQT predicting the PIQ and FSIQ demonstrated a race main effect, while most of the AQE AIs exhibited some type of interaction occurring between race of trainee and his AQE AI. These findings agree with Guinn, Tupes, and Alley who report several

significant interactions between aptitude test and demographic variables.

Significant differences between black and white scores occurred on the AFQT, the four AQE AIs, and the three WAIS IQs. Blacks performed uniformly lower on both tests of general mental ability and selection tests. Before any conclusions on the abilities of blacks can be made, however, a determination of exactly what these test scores are revealing is necessary. An examination of the WAIS subtests demonstrates that blacks scored significantly lower on all subtests, except Digit Span and Digit Symbol where blacks and whites are statistically equivalent. These two areas are unique in that they are perhaps the least related of all the WAIS subtests to the skills acquired through the formal schooling of the subject. Digit Span requires the subject to repeat series of numbers forward and backward, while Digit Symbol is a test of how accurately the subject can match numbers and unfamiliar symbols. Memory, attention, and visual perception are three essential abilities measured by these two subtests.

Guinn, Tupes, and Alley have categorized perceptual and psychomotor skills of this type as situational abilities, as opposed to prior knowledge abilities, which are a function of an individual's educational background. They found, and our WAIS data corroborated, that prior knowledge items produce greater racial differences in test scores than situational items. Perhaps there is an over-abundance of prior knowledge items within the AQE subtests, which are responsible,

in part, for the differences between blacks and whites in AI scores. Future revisions of selection tests, especially the ASVAB now coming into use, should include an investigation into the degree of importance assigned to prior knowledge and situational variables in the construction of the test. Such an endeavor may discover a control for qualitative disparities in educational level of black and white recruits which may be masking the identification of true ability in these tests.

In addition to item specificity, a valid interpretation of the black-white score differences involves a close examination of the test interrelationships. The correlation matrices and the regression R^2s show a much closer relationship between the selection tests and the WAIS for whites than for blacks. The consistently high intercorrelations of the white sample indicate a good deal of common variance shared between tests. The much lower intercorrelations of the black sample evince that for them the selection tests and the WAIS are measuring largely divergent factors. Further research into the internal characteristics of these tests is necessary to determine the cause of this disparity between whites and blacks in test commonality.

Two possible areas of inquiry are the literacy factor and the administrative format of the AQE and WAIS. The AQE is a paper and pencil test requiring the subject to read the instructions and test items. WAIS instructions and test items, on the other hand, are presented orally by the test administrator. Only the

Vocabulary subtest involves the subject reading anything. He must read and define vocabulary words from a card, but simultaneously the administrator is pronouncing the words. Thus, the subject is not solely dependent upon his ability to read. A study measuring the differential literacy requirement of the two tests appears in order.

The second area involves the tester. The WAIS is administered individually by professionally trained testers. The AQE, however, is a group test administered and scored by recruiting testers who have had minimal formal training in testing. Recruiter effects in the distribution of scores are readily apparent. For example, the clustering of scores just above AQE AI cutoffs for career field entry was found in this study and others conducted by this division. A rescoring of the AQEs administered to our subjects, or even better, the administration of the AQE to a selected sample of airmen in a situation more conducive to test taking than a recruiter's office, may produce more valid AQE scores.

In conclusion, our data reveals that black Air Force recruits score significantly lower than white recruits on the AFQT, AQE, and WAIS. What this means in terms of the true abilities that blacks and whites are bringing into the Air Force involves the identification and control of variables in these tests, such as literacy and item specificity, that are masking a valid measurement of ability. Subsequent research should explore the test format, administrative environment, and item pools of Air Force selection tests to determine exactly what is being measured by

these tests.

Finally, we would like to comment on racial bias. We have purposefully couched our results in terms of racial differences in scores rather than explaining them in terms of racial bias. Any interpretation of the cultural fairness of the AQE based upon these data must be tempered by the fact that the criterion test, the WAIS, was an experimental test for our purposes, rather than an operational instrument in the Air Force selection process. According to Thorndike (1971), it is the use to which the test is put, rather than the test itself, which defines the fairness of the test. In other words, the fairness of a selection test is determined by its application in the selection process, not its internal characteristics. Shore and Marion put forth their conclusions on the fairness of the AQE based upon its relationship to the SKT, an operational test employed b the Air Force to measure one factor in promotion. We can only present our data in relation to the WAIS, which we have utilized as a baseline for measuring general mental ability. Hopefully, some clarification of the racial differences in the AFQT and AQE will accrue from this study. But we must leave a definitive interpretation of racial bias in selection testing to other investigations.

References

Guinn, Nancy, Tupes, E. C., and Alley, W. E. Demographic differences in aptitude test performance. AFHRL-TR-70-15. Lackland AFB, Tex: Personnel Research Division, Air Force Human Resources Laboratory, May 1970.

Shore, C. W., and Marion, R. Suitability of using common selection test standards for Negro and white airmen. AFHRL-TR-72-53. Lackland AFB, Tex: Personnel Research Division, Air Force Human Resources Laboratory, May 1972.

Thorndike, R. L. Concepts of culture-fairness. Journal of Educational Measurement, 1971, 8, 63-70.

Table 1. Means and Standard Deviations for Tests

Test	Black Sample (N = 100) Mean	SD	White Sample (N = 100) Mean	SD	t
AFQT	41.68	13.83	64.98	20.67	9.32
AQE-A	44.55	16.82	56.01	21.47	4.18
AQE-E	45.25	18.56	63.65	20.06	6.70
AQE-G	48.26	16.72	60.90	19.41	4.91
AQE-M	44.20	18.56	62.75	18.95	6.96
VIQ	96.74	8.78	104.99	10.47	6.02
PIQ	97.41	9.65	109.30	10.57	8.26
FSIQ	96.72	8.33	107.17	9.90	8.04
Information	8.87	2.08	10.77	2.34	6.03
Comprehension	9.21	2.67	10.83	2.74	4.21
Arithmetic	8.68	2.09	10.70	2.48	6.20
Similarities	9.61	2.39	10.46	2.59	2.39
Digit Span	9.19	2.76	9.67	2.46	1.29
Vocabulary	8.74	1.84	10.24	2.15	5.26
Digit Symbol	9.76	2.16	10.33	2.17	1.85
Picture Completion	9.69	1.75	11.46	2.38	5.96
Block Design	9.21	2.27	12.22	2.66	8.58
Picture Arrangement	9.97	2.11	11.12	2.45	3.53
Object Assembly	9.01	2.31	11.73	3.08	7.03

With 198 df t = 1.98 at .05 level.
t = 2.61 at .01 level.

Table 2. Intercorrelations of WAIS, AFQT, and AQE Scores for the Black Sample (N=100)

		WAIS				AQE			
		VIQ	PIQ	FSIQ	AFQT	A	E	G	M
	VIQ		.46	.88	.38	.27	.26	.23	.10
WAIS	PIQ			.83	.44	.19	.26	.29	.23
	FSIQ				.48	.30	.30	.30	.19
	AFQT					.26	.37	.36	.30
	A						.31	.62	.29
AQE	E							.55	.40
	G								.63
	M								

Table 3. Intercorrelations of WAIS, AFQT, and AQE Scores for the White Sample (N=100)

		WAIS				AQE			
		VIQ	PIQ	FSIQ	AFQT	A	E	G	M
WAIS	VIQ		.54	.91	.64	.60	.69	.77	.45
	PIQ			.84	.62	.28	.60	.44	.50
	FSIQ				.71	.52	.74	.71	.53
	AFQT					.51	.75	.64	.67
AQE	A						.56	.71	.34
	E							.73	.69
	G								.56
	M								

Table 4. Regression Design

Regression Models

Model 1: $WAIS\ IQ' = w_1B + w_2W + w_3(B \times Sel\ Test) + w_4(W \times Sel\ Test)$

Model 2: $WAIS\ IQ' = w_1B + w_2W + w_3(Sel\ Test)$

Model 3: $WAIS\ IQ' = w(Sel\ Test)$

F - Tests

$$F = \frac{(R_F^2 - R_R^2)/df_1}{(1 - R_1^2)/df_2}$$

	Full Model		Restricted Model	Testing for:
Analysis I	Model 1	vs	Model 3	Racial Difference
Analysis II	Model 1	vs	Model 2	Slope Difference
Analysis III	Model 2	vs	Model 3	Intercept Difference

Definitions

WAISIQ = VIQ, PIQ, or FSIQ score
w = Variable weight
B = 1 if black, 0 if not black
W = 1 if white, 0 if not white
Sel Test = AFQT or AQE AI percentile score

R_F^2 = Squared multiple correlation of full model

R_R^2 = Squared multiple correlation of restricted model

R_1^2 = Squared multiple correlation of model 1

df_1 = Independent vectors in full model minus independent vectors in restricted model

df_2 = Elements in vectors minus independent vectors in model 1

Table 5. Summary of Regression Analyses

F tests for Presence of Racial Difference (slope or intercept differences)[1]

Model 1 vs Model 3

Criterion	Predictor	R_F^2	R_R^2	df_1	df_2	F	prob
VIQ	AFQT	.40742	.40058	2	196	1.13	NS
VIQ	AQE-A	.35580	.27294	2	196	12.61	.001
VIQ	AQE-E	.41354	.34079	2	196	12.16	.001
VIQ	AQE-G	.46722	.36258	2	196	19.25	.001
VIQ	AQE-M	.25999	.17376	2	196	11.55	.001
PIQ	AFQT	.47611	.44930	2	196	5.01*	.01
PIQ	AQE-A	.29991	.11748	2	196	25.54*	.001
PIQ	AQE-E	.42403	.31719	2	196	18.18	.001
PIQ	AQE-G	.36390	.22263	2	196	21.76*	.001
PIQ	AQE-M	.37527	.26173	2	196	17.81	.001
FSIQ	AFQT	.54021	.52540	2	196	3.16*	.05
FSIQ	AQE-A	.38758	.24062	2	196	23.52*	.001
FSIQ	AQE-E	.51365	.41136	2	196	13.74	.001
FSIQ	AQE-G	.49822	.36716	2	196	25.60	.001
FSIQ	AQE-M	.38220	.26353	2	196	18.82	.001

[1]Note. Starred (*) F tests denote race main effects (intercept differences but no slope difference). All other significant differences were shown to be interaction effects (both slope and intercept differences).

Does the USAF Officer Biographical Inventory Portion of the AFOQT Inadvertently Measure Authoritarian Personality?

a paper for presentation at the 15th Annual Conference of the
Military Testing Association
28 October - 2 November, 1973

by

Lt Col Charles W. Haney
Chief, College Curriculum Division
Air Force Reserve Officers Training Corps
Maxwell AFB, Alabama

I. INTRODUCTION

In the present all-volunteer armed forces environment, it is particularly crucial to our national interests that we maintain an adequate and effective officer corps. While academies and officer training schools provide significant inputs, Reserve Officer Training Corps (ROTC) commissioning programs provide the greatest single source of newly commissioned officers for the armed forces.

ROTC programs have experienced a significant decline in enrollments in recent years. There are a number of factors to which this decline can be attributed; the zero-draft, anti-militarism and apathy toward the military, and the relatively large scale eradication of compulsory ROTC programs on America's campuses. In an effort to offset this decline, the ROTC programs have expended considerable efforts to maintain viable programs on our nation's campuses.

Alternatives to campus-based commissioning programs have not been particularly appealing even to many who have been less than sympathetic toward the military. The leavening input, via ROTC, of young officers fresh from the humanizing and liberalizing influences of the college campus environment tends to offset, according to many, the relatively rigid and authoritarian structure of the military environment. (Karsten et al, 1971)

II. BACKGROUND

We at Air Force ROTC have been concerned with providing a respectable campus-based procurement source for young officers of high quality. In recent years the emphasis of our program has been placed on attracting and motivating qualified students to pursue our pre-commissioning program. We have been particularly sensitive to certain assertions that ROTC tends to attract students with traits related to militarism, authoritarianism, and other behavior patterns which might be considered contrary to the traditional egalitarian atmosphere of the college campus. (Gage, 1964; French and Ernst, 1955; Lucas, 1967)

A point of concern was whether or not the Air Force Officer Qualifying Test (AFOQT), one of several selection criteria for cadets seeking entry into the advanced AFROTC course, or seeking scholarship status, might be inadvertently functioning to screen out the more egalitarian students.

One portion of the test, the Officer Biographical inventory (OBI), became suspect as a result of a study conducted by the AFROTC Education Evaluation and Research Division in the summer of 1970. (Gordon, 1970) In this study it was found that a relatively low correlation existed between grade point averages (GPA's) and the Officer Quality Composite (OQC) scores of the AFOQT for 468 cadets attending AFROTC Field Training. Table I contains these correlations.

TABLE I

Correlation of OQC Components With GPA
(Gordon's Study, 1970) N = 468

Correlates	r
Overall OQC	.22
Verbal	.20
Quantitative	.20
Officer Biographical Inventory	-.06

The relatively low coefficients of correlation between GPA's and the verbal and quantitative portions of the OQC notwithstanding, Gordon reasoned that the lack of correlation between GPA and the OBI reflected cause for concern. He further analyzed relationships between the OBI and overall OQC, Verbal and Quantitative components, GPA, and the AFROTC Field Training Rating. Table 2 contains these correlations.

TABLE 2

Correlation of OBI With Overall AFOQC, Verbal, and Quantitative Components, GPA, and AFROTC Field Training Rating (Gordon's Study, 1970) N = 468

Correlates	r
Overall AFOQC	.26
Verbal	-.17
Quantitative	-.24
GPA	-.06
Field Training Rating	.14

These findings led Gordon to suspect that the OBI might be dysfunctional to the overall OQC. It might, for instance, be detracting from the purported relationship between the OQC and overall academic achievement. It might therefore be detracting from the intended purpose of the OQC which is "to evaluate aptitudes which are important for commissioned officer performance and success." (Air Force Manual 35-8, 1969)

The OBI component of the OQC is purported to consist of "items pertaining to past experiences, preferences, and personality characteristics known to be related to success in officer training." (Miller, 1969) This fits a general

description of measures of attitude. There are many varieties of attitude, however, which can be treated as independent variables. The purported purpose of the OBI does not specifically identify any particular variety of attitude. Further study was needed, therefore, to shed additional light on the true function of the OBI. This prompted the present study to investigate the relationship between the OBI and known measures of authoritarian personality.

Since other researchers had speculated that military officer training programs tend to attract those with "militaristic" and "authoritarian" traits of personality (Lammers, 1965; Lucas, 1967), it seemed appropriate to choose a trait such as "authoritarianism" for comparison to the OBI. It was reasoned that useful information could be obtained, through baseline research procedures, by correlating OBI raw scores with scores obtained from measures of authoritarian personality. If a significant relationship was found to exist between these sets of data, then a relationship might be inferred to exist between the OBI and measures of authoritarian personality.

III. PURPOSE OF THE STUDY

The purpose of this study was to determine if a significant relationship could be inferred to exist between the Officer Biographical Inventory (OBI) component of the Officer Quality Composite (O_QC) of the Air Force Officer Qualifying Test (AFOQT) and measures of authoritarian personality. The

hypothesized relationship, if confirmed to exist, might have been the basis for criticism not only of the AFOQT but of the Air Force ROTC Program selection and commissioning process which the AFOQT directly affects.

IV. PROCEDURE

The California F-scale was selected as a suitable instrument to measure authoritarian personality for the purpose of this study. Despite some criticism of the F-scale, it has been widely accepted as a valid measure of authoritarian traits. Three versions of the F-scale were adopted for use in this study. These were the Form 78, and Forms 40 and 45. These were taken virtually intact from the pages of The Authoritarian Personality, Adorno et al., (1950). An additional scale was developed and validated by the principal investigator and Dr. Earl Jones, Texas A&M Univ., to act merely as a stabilization cross-check of the F-scales. Adorno's Form 78 is referred to hereinafter as F-1. Forms 40 and 45 were combined into one scale and are referred to hereinafter as F-2. The special cross-check scale is referred to as the JOHAN (Jones and Haney) scale.

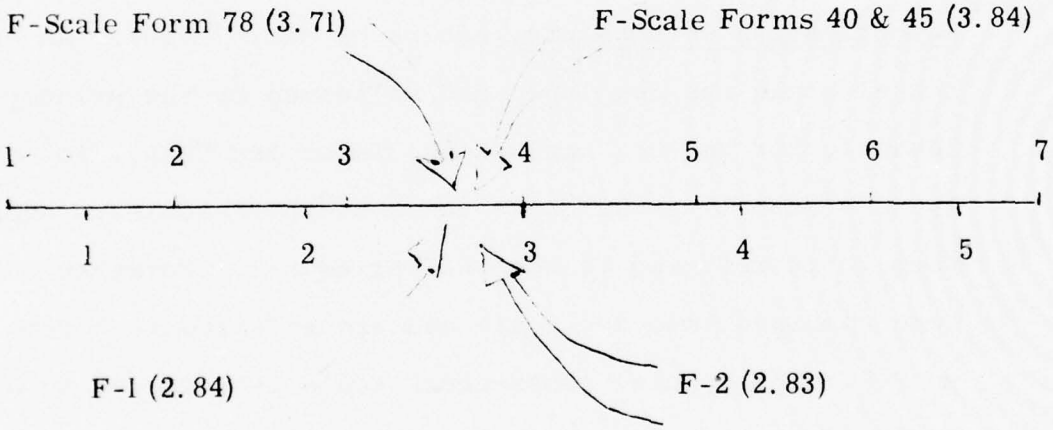
The Air Force ROTC cadet population at Texas A&M University (April, 1971) was selected as subjects for this study because of the relative availability of AFOQT data on this group. Scales F-1, F-2, and JOHAN were administered to the Texas A&M cadets and this yielded 391 usable samples. At the time these scales were administered, this number represented

approximately 60 percent of the Texas A&M AFROTC cadets who had taken the AFOQT. This was deemed a sufficient number of respondents to conduct a valid and significant study.

An analysis of F-1 and F-2 scores from the sample population revealed that the mean scores obtained were very close to the norms cited by Adorno et al., (1950). Figure 1 illustrates this comparison. The F-1 and F-2 scales adapted to this study used five point Likert-type scales while the Adorno scales were seven point scales. Nevertheless the comparisons were quite favorable.

FIGURE 1

Comparison of Norms for F-Scales



OBI raw scores on each of the respondents were extracted from data available at AFROTC headquarters at Maxwell AFB. The OBI mean for the sample population was 37. This also compared quite favorably with the Air Force-wide OBI raw score mean estimated to be 36 by the Air Force Human Resources Laboratory. These sets of data, the OBI scores, F-1, F-2, and JOHAN scores, became the basis of comparison for the purpose of this study.

As a first step in testing the relationship between the various measures used in this study, the following null hypotheses were formulated:

H_{o_1} There would be no significant relationship between OBI raw scores and F-1 raw scores.

H_{o_2} There would be no significant relationship between OBI raw scores and F-2 raw scores.

H_{o_3} There would be no significant relationship between OBI raw scores and JOHAN raw scores.

H_{o_4} There would be no significant relationship between F-1 raw scores and F-2 raw scores.

H_{o_5} There would be no significant relationship between F-1 raw scores and JOHAN raw scores.

H_{o_6} There would be no significant relationship between F-2 raw scores and JOHAN raw scores.

V. FINDINGS

The OBI raw scores were correlated with scores from each of the scales used to measure authoritarian traits and with the JOHAN scale. Pearson coefficients were obtained as depicted in Table 3. The only significant relationships found to exist were those between the F-1, F-2, and JOHAN scales.

TABLE 3

Coefficients of Correlation:

Entire Population

Scale	OBI	F-1	F-2	JOHAN
OBI		-.06	-.03	.06
F-1			.78*	.46*
F-2				.61*

* significant at .05

On the basis of these data, the investigator failed to reject the following null hypotheses: H_{o_1}, H_{o_2}, and H_{o_3}. The investigator did reject H_{o_4}, H_{o_5}, and H_{o_6}. The investigator sought to determine if this lack of correlation between the OBI and measures of authoritarianism was a general characteristic of the sample population or was possibly a function of the combined class groupings. Realizing, for instance, that lower OBI scores would have been screened out in the process of selecting cadets into the Professional Officer Course (POC), the advanced level course offered during the last two years of AFROTC, and that greater authoritarianism

was probably to be found among the lower classmen, the investigator decided that further computations were necessary in order to treat each class as a separate entity. Correlation coefficients were therefore computed for each class level adhering to the same null hypothesis in each case. Tables 4 through 7 show the correlation coefficients obtained. Again, in each case, the investigator failed to reject the first three null hypotheses and rejected the last three. The results were quite similar to the correlation coefficients obtained from the combined group.

TABLE 4

Coefficients of Correlation:

Freshmen

Scale	OBI	F-1	F-2	JOHAN
OBI		-.03	-.04	-.11
F-1			.81*	.43*
F-2				.59*
JOHAN				

* significant at .05

TABLE 5

Coefficients of Correlation:

Sophomores

Scale	OBI	F-1	F-2	JOHAN
OBI		-.09	-.06	-.11
F-1			.82*	.59*
F-2				.66*
JOHAN				

* significant at .05

TABLE 6

Coefficients of Correlation:

Juniors

Scale	OBI	F-1	F-2	JOHAN
OBI		.13	.07	.07
F-1			.78*	.57*
F-2				.69*
JOHAN				

* significant at .05

TABLE 7

Coefficients of Correlation:

Seniors

Scale	OBI	F-1	F-2	JOHAN
OBI		.05	.03	.13
F-1			.77*	.47*
JOHAN				.63*

* significant at .05

Conclusions

The principal conclusions drawn from this study are as follows:

1. The raw scores obtained from the measures of authoritarianism used in this study did not correlate significantly to corresponding OBI raw scores. This lack of correlation was found to exist when treating data from the entire sample population as well as when treating data from separate classes within the sample population.
2. The separate measures of authoritarianism did correlate significantly to one another when treating data from the entire sample population as well as when treating data from separate classes within the sample population.

The first conclusion was drawn by the investigator after having failed to reject the null hypotheses H_{o_1}, H_{o_2}, and H_{o_3}, stated earlier in this study, at the .05 level of confidence. The same conclusion was again drawn after having failed to reject these same null hypotheses at the .05 level of confidence when treating similar sets of data for each of the separate classes within the sample population, i.e., the freshman, sophomore, junior, and senior classes. The second conclusion was drawn after having rejected the null hypotheses H_{o_4}, H_{o_5}, and H_{o_6}, at the .05 level of confidence, also stated earlier in this study. Again data were treated for not only the entire sample population but for the separate classes within the sample population.

From these conclusions, the investigator drew certain inferences. The first of these was that the OBI could not be considered to be a measure of authoritarian personality traits. This inference was based primarily on the assumed validity of the F-scale as a measure of authoritarianism, and the lack of correlation between raw scores of the F-scale and the OBI.

The investigator then inferred that if the OBI did not measure authoritarianism, it could not be used as a means by which to predict authoritarianism. It was further inferred that the OBI probably would not predict many of the correlates of authoritarianism purported to be measured in some degree by the F-scales such as ethnocentrism, prejudice, and anti-semitism (Adorno et al., 1950).

The investigator then inferred that if the OBI neither measured nor predicted authoritarianism, then the OBI, as an instrument of measurement and prediction, has no relevance in any meaningful discussion of authoritarianism. The same might further be true in any meaningful discussion of the correlates of authoritarianism.

These inferences were not meant to imply that the OBI could not be used to discuss some other construct in a meaningful way. There are a number of traits which are widely accepted as indicative of success in Air Force officers. Motivation and interest factors undoubtedly play a very important role in the success of officer candidates in training or active duty officers. Other traits include general leadership qualities such as intelligence, initiative, resourcefulness, creativity, loyalty, dedication, and patriotism, to name only a few. Other personality traits, as reflected by various attitude propensities, are also undoubtedly important.

VI. RECOMMENDATIONS

This study was not designed to, nor did it address the question of the overall validity of the OBI. It was not meant to explore the OBI's ability to measure all of the various traits which may be predictive of success in officer training and performance. Although there are undoubtedly many measures of personality traits with which the OBI

could be favorably compared, there are few such studies available in the literature.

It is therefore recommended that further studies be accomplished in order to compare the OBI to other measures of personality. This would shed new light on the overall validity of the OBI as an instrument designed to measure interests, preferences, and propensities of modern day youth.

SELECTED REFERENCES

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J. and Sanford, R. N. The Authoritarian Personality. New York: Harper and Row, 1950.

French, E. G. and Ernst, R. R. The relationship between authoritarianism and the acceptance of military ideology. Journal of Personality, 1955, 24, 181-191.

Gage, R. W. Patriotism and military discipline as a function of the degree of military training. Journal of Social Psychology, 1964, 44, 101-111.

Gordon, M. M., Jr. What does the AFOQT score predict? Unpublished paper. Maxwell Air Force Base, Alabama: Evaluation and Research Division, Education Directorate, Air Force ROTC, 1971.

Haney, C. W. A study of the relationship between selected measures of authoritarian personality and the Officer Biographical Inventory portion of the Air Force Officer Qualifying Test. (Doctoral dissertation, Texas A&M University) Ann Arbor, Mich: University Microfilms, 1972.

Karsten, P. et al. ROTC Mylai and the volunteer army. Foreign Policy, 1971, 2, 135-160.

Lammers, C. J. Midshipmen and candidate reserve officers at the Royal Netherlands Naval College: a comparative study of the socialization process. Sociologia Neerlandia, 1965, II, 98-122. Cited by Karsten et al., ROTC, Mylai and the volunteer army. Foreign Policy, 1971, 2, 135-160.

Lucas, W. A. The American lieutenant: An empirical investigation of normative theories of civil-military relations. Unpublished doctoral thesis, University of North Carolina, 1967.

Miller, R. E. Interpretation and utilization of scores on the AFOQT. AF-HRL-TR-69-103. Lackland Air Force Base, Texas: Air Force Human Resources Laboratory, Personnel Research Division, May, 1969.

United States Department of the Air Force. Air Force Manual 35-8. Air Force Military Personnel Testing Manual. Washington, D. C., 1968.

HUMAN RELATIONS EVALUATION

Ms. A. M. KELLEHER
Personnel Psychologist
United States Coast Guard Training Center, Governors Island, New York 10004

PURPOSE

To study the effect of the Human Relations course participation on the leadership factors of consideration and structure was the purpose of this study.

INTRODUCTION

Since the four member Human Relations Team has been conducting fifteen hour human relations classes at TRACEN, N. Y., two thousand nine hundred and five TRACEN personnel, seventy-three Base N. Y. personnel, and one hundred and forty-one other CG personnel have received training in this area. All ranks and duties have been included among the participants.

Questionnaires on the topics covered in class which attempted to measure very grossly the participants' attitudes were utilized throughout the six month period of June - December 1972. These questionnaires were tabulated and summarized by the TRACEN Psychologist and indicated certain trends within ranks on beliefs about the amount of racial descrimination at the TRACEN and in the Coast Guard, beliefs about leadership, and beliefs about military problems/benefits verses civilian problems/benefits. The main results were that Chiefs and Captains had similar responses and PO3s and Lieutenant Junior Grades responded most similarly. It was theorized that a man's place on the totem pole with its commensurate authority and respect within the enlisted and officer structure seemed the best way to predict pattern of belief. The other main conclusion was that all felt leadership was a more severe problem than race and also the answer to better human relations within the Coast Guard.

However, since these questionnaires were short, hard to quantify, and frequently reworded as additions/deletions were made to the course the TRACEN Psychologist considered it tenuous evidence for conclusions about Coast Guard personnel. Also, the main objective, to evaluate the Human Relations course was not being achieved.

Consideration of the results; however, directed the command to the area of leadership. This was one topic consistently mentioned by the Human Relations participants. It was on this topic that the Human Relations course effectiveness could perhaps be evaluated.

After a survey of the literature, the TRACEN Psychologist decided to utilize the Leadership Opinion Questionnaire published by Science Research Associates, Inc. This questionnaire has been used to evaluate human relations training (Harris & Fleishman, 1955) and it is designed to measure two main factors of leadership, consideration and structure.

> Consideration (C). Reflects the extent to which an individual is likely to have job relationships with his subordinates

characterized by mutual trust, repect for their ideas, consideration of their feelings, and a certain warmth between himself and them. A high score is indicative of a climate of good rapport and two-way communication. A low score indicates the individual is likely to be more impersonal in his relations with group members.

Structure (S). Reflects the extent to which an individual is likely to define and structure his own role and those of his subordinates toward goal attainment. A high score on this dimension characterizes individuals who play a very active role in directing group activities through planning, communicating, information, scheduling, criticizing, trying out new ideas, and so forth. A low score characterizes individuals who are likely to be relatively inactive in giving direction in these ways.

Therefore, the problem had been defined, leadership; the measurement tool procured, Leadership Opinion Questionnaire; and the subjects were selected, 30 CG personnel at TRACEN, GINY - 15 graduates of the Human Relations course mathched by rank to 15 non-attendees.

It was theorized that Consideration would be the factor most likely to be affected by Human Relations Training, although the emphasis on direct communication in the course should also influence the Structure factor.

METHOD

The questionnaire was test run on a Human Relations class of nine Commanders to evaluate its ease and pitfalls of administration and tabulation. Also the Commanders were asked to evaluate it. The questionnaire was evaluated as a usable and appropriate measure for the Human Relations course by the TRACEN Psychologist.

The Executive Officer of the TRACEN chose the 30 subjects and delivered a questionnaire and coverletter to them. The questionnaire was to be anonymous and returned within two days to the XO. The only information they were to give was whether they had attended Human Relations course or not.

The recipients were matched by rank.

Group 1 (did not attend)	Group 2 (attended)
1 Commander	1 Commander
1 Lieutenant Commander	1 Lieutenant Commander
1 Lieutenant	1 Lieutenant
1 Lieutenant Junior Grade	1 Lieutenant Junior Grade
4 Chief Warrant Officers	4 Chief Warrant Officers
7 Chief Petty Officers	7 Chief Petty Officers
15	15

Twenty-nine questionnaires were returned. Since they were anonymous, the only information available is that the gentleman who did not return a questionnaire did not attend the Human Relations course. The results were tabulated into raw scores of the factors of Consideration and Structure for each man. These raw scores were then converted into percentile scores using the Norms Table provided by the Science Research Associates company for the Questionnaire (Table 1). The comparison norms selected were the General Supervisory Personnel norms since the groups were combined of different ranks, and more specific norms, e.g. Executive norms would be appropriate for only a few members of the test groups.

RESULTS

The mean scores indicated that the 14 graduates of the course scored higher on both Consideration (61.57) and Structure (58.07) than the non-attendees (53 on onsideration and 51.2 on Structure).

A t test was deemed most appropriate to measure the significance or non-significance of these differences. Two t tests were run:

1. Group 1 versus Group 2 compared on Consideration; the t test yielded a value of 6.47. With 28 degrees of freedom, this value is significant beyond the .01 level.

2. Group 1 versus Group 2 compared on Structure; the t test yielded a value of 9.54. With 28 degrees of freedom, this is significant beyond the .01 level.

Comparisons between Consideration and Structure within the two groups were also conducted:

1. Group 1, Consideration versus Structure; the t test yielded a value of 1.8. With 14 degrees of freedom, this is not significant.

2. Group 2, Consideration versus Structure; the t test yielded a value of 3.56. With 13 degrees of freedom, this is significant beyond the .01 level. The Consideration Score is significantly higher for Group 2.

All mean scores , however, were above the 50th percentile (Table 1).

CONCLUSIONS

Attendance at the Human Relations course results in significantly higher scores on leadership factors of Consideration and Structure as measured by the Leadership Opinion Questionnaire given to two matched groups of attendees and non-attendees. The graduate scores higher than a man of similar rank and expearience who did not attend.

Consideration appears to receive higher significant scores than Structure among the graduates, and this factor may characterize the Coast Guard Human Relations graduate leader.

TABLE 1

RAW SCORES AND PERCENTILE NORMS FOR BOTH GROUPS

	Group 1 (did not attend)				Group 2 (attended)			
	Consideration		Structure		Consideration		Structure	
	raw score	percentile	raw score	percentile	raw score	percentile	raw score	percentile
1.	56	65	53	65	56	65	52	60
2.	55	60	46	28	49	27	47	31
3.	45	12	57	80	51	40	43	17
4.	58	75	55	75	69	98	57	80
5.	62	90	44	20	49	28	50	50
6.	49	28	56	78	57	69	56	77
7.	60	85	39	08	66	95	63	95
8.	55	60	54	69	50	31	40	08
9.	50	31	41	10	52	45	52	60
10.	63	93	50	50	60	90	55	75
11.	59	80	58	85	63	91	52	60
12.	36	01	47	31	48	25	53	65
13.	51	40	44	10	55	60	55	75
14.	55	60	60	90	71	99	52	60
15.	46	15	54	69				
	$\bar{X}$ = 53		$\bar{X}$ = 51.2		$\bar{X}$ = 61.57		$\bar{X}$ = 58.07	
	n = 15				n = 14			

REFERENCES

Fleishman, E.A. Manual for Leadership Opinion Questionnaire. Chicago: Science Research Associates, Inc., 1969.

Harris, E.F., and Fleishman, E.A. Human Relations Training and the Stability of Leadership Patterns, Journal of Applied Psychology, 1955, 39, 20-25.

Centralized Selection System For Air Force Pilots

by
Melvin S. Majesty, Lt Col, USAF

Pilot training is the most costly type of individual training given by any of the services. In fact, it is probably the most expensive educational process in the world. In recent years, the bill for pilot training for the military services amounted to over two billion dollars a year.

The cost of pilot training depends upon which costs are included in or excluded (e.g. aircraft costs) from the calculation. The costs per graduate from Undergraduate Pilot Training (UPT) is about $90,000. The initial flight training of an Air Force pilot takes almost a year. The program includes flying training (using T-37 and T-38 jet aircraft) of sufficient scope to produce the necessary understanding of the principles and proficiency requirements associated with flying high speed jet aircraft. The UPT program produces a graduate who is capable of transition into any aircraft in the Air Force inventory. He must now be trained to fly an operational aircraft. Upon graduation from UPT, the pilot is assigned to an operational command for training in a specific aircraft. In the operational flying training program, e.g. Combat Crew Training, the specific procedures unique to an individual aerospace system and missions are emphasized.

New Centralized Program

Recently the Air Training Command completed a mission analysis of pilot training for the 1975 through 1990 time period (Mission Analysis Study Group, 1972). This mission analysis included the interface of the screening and selection process with the future Undergraduate Pilot Training (UPT) system. Properly designed screening and selection was considered a critical component for the future UPT system. It was concluded that a "ten percent attrition level is a realistic goal if centralized selection is employed and training improvements are implemented."

The initial step in the centralized selection program was the establishment of a light plane (military version of the Cessna Model 172) flight screening program as part of the officers training school. Officers who had previously completed a USAF sponsored light plane program would be excused from the Flight Screening Program (FSP). All other pilot candidates would complete FSP prior to UPT.

Along with the FSP came the centralized Class I physical examination and a program to test and validate various ground base screening devices for operational use. The centralized selection program makes the use of individually administered tests feasible. It is anticipated that a ground base screening system will emerge to replace or partially replace the initial light plane screening program (FSP).

The ground based devices will be used to predict success in the light plane FSP program as well as success in the jet trainer's (T-37/T-38). At the same time it will be possible to determine the extent to which the FSP will predict success in jet training.

The development, test and validation of a new ground based pilot screening system is the first major effort on pilot selection to be undertaken by

Air Force since World War II. Over the years the Air Force has continued to update and revise the operational system and the selection procedures. Selective studies have been accomplished in the area of pilot training. However, this is the first time a total new system is being investigated as a complete entity.

Selection Procedure

Although there are several sources of entry into pilot training, (Air Force Academy, AFROTC, etc) the largest single source of both entry and washout in the Undergraduate Pilot Training (UPT) program is School of Military Sciences, Officer. Most of the entries into the officer training school (School of Military Sciences, Officer) come through direct recruitment. The procedure begins with an interview in the Recruitment Office to determine whether the candidate meets the minimum requirements, e.g. college graduate, citizen, and he must be able to enter UPT prior to age 27½. If he is a potential candidate, he is given the Air Force Officer Qualifying Test (AFOQT). Upon scoring of the AFOQT a determination is made as to whether he is fully qualified or meets the minimum standards on the AFOQT (Figure 1). A fully qualified candidate completes the application form and takes a flight physical examination. These are sent to the Flight Surgeon's Office for review. He is then mailed a class entry card which specifies the date he is to begin training. He must sign and return this card to the Air Force. If his scores on the AFOQT are above the minimum standards but not above the fully qualified standards, the procedure is a little different. The application form is first sent to the Selection Board. If the Selection Board accepts the applicant, the flight physical examination is given and the results sent to the Flight Surgeon's Office for review. A class entry card is mailed to the applicant for signature.

Potential candidates can be lost anywhere in the selection process.

Selection Process

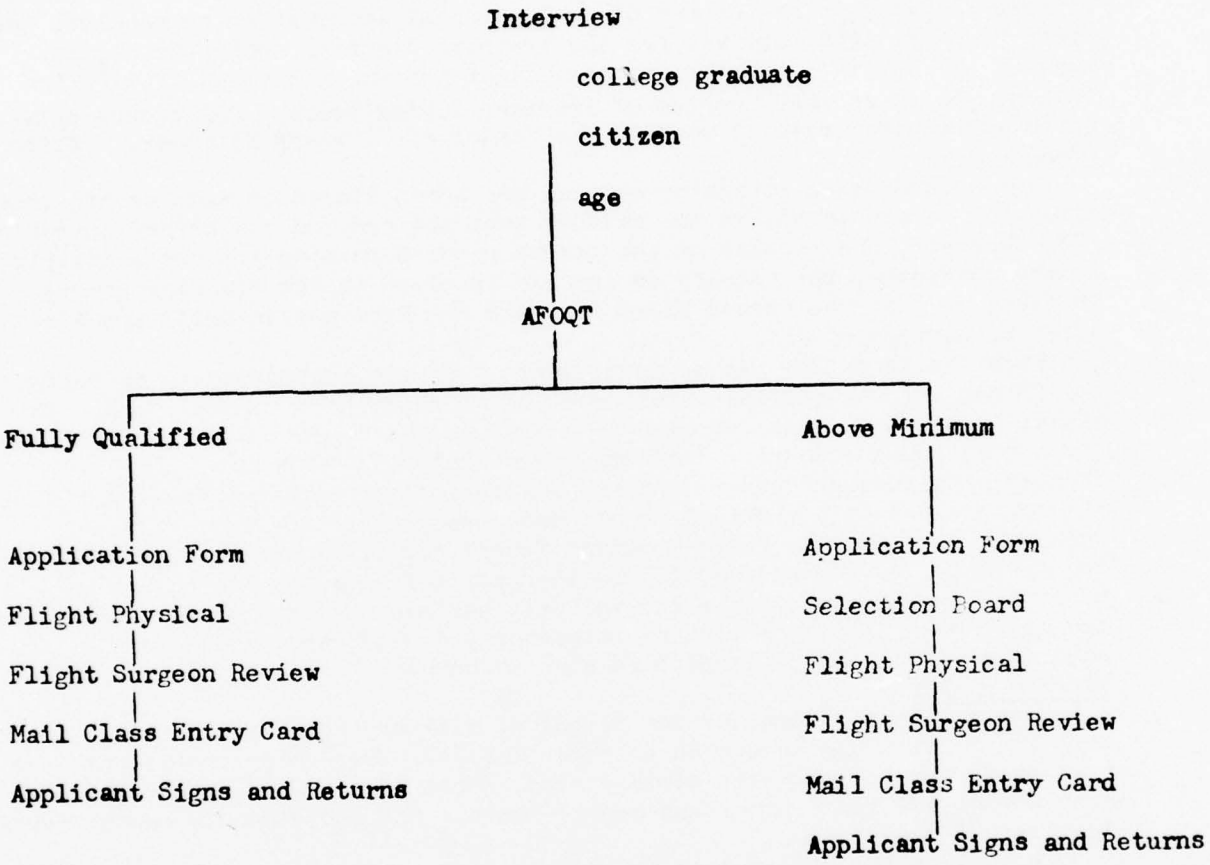


Figure 1

Concept of Operation

Several selection techniques will continue to be use during the validation of ground based screening devices. If the potential candidate does not satisfy the minimum requirements as determined through the interview, he cannot apply for pilot training. If he does not qualify on the AFOQT, he cannot submit an application. If his application has to be acted upon by the Selection Board, he must be favorably considered by the Board before he can take the flight physical. He may be disqualified on the basis of this flight physical or the Class I physical examination he receives the first few days after arrival for the training program. After he begins training he may be eliminated in the Flight Screening Program (FSP) which is a six to seven week program of fourteen flying hours. FSP occurs between the third and fourth training week in the School of Military Sciences, Officer (SMSO).

The ground based screening devices are administered as soon as possible after the candidate enters the officer training program and before he begins FSP. However, the results of the ground based screening are not available to the candidate, the faculty or any one involved in the training program. In fact, most of the ground based test are not even scored until sometime after completion of FSP.

Emphasis is on the use of instruments that are considered to be "off-the-shelf" because the initial development has already been accomplished. This minimizes the lead time required to place the instrument in operational use. Assessment areas encompass aptitude, psychomotor performance, [illegible] chamber, attitude and career intent, motivation and career indoctrination. The optimum use of a light plane screening device will be investigated. Later, when resources permit, Basic Survival School may be tested as a pre-UPT screening device. New techniques will be incorporated into the operational selection system as soon as possible after validity has been demonstrated. Concurrently, the light plane flying screening program may be reduced, or altered to incorporate more efficient in-flight screening techniques.

Program Design

Data collection began at the School of Military Sciences, Officer (SMSO) with the first class scheduled to enter the FSP. Screening instruments are both group and individually administered. Most of the group administered instruments are timed paper and pencil tests. The individually administered test tend to be untimed.

In view of the differences in the types of tests and test availability the program was organized into three phases:

Phase I	May 1973 - November 1973
Phase II	November 1973 - March 1974
Phase III	March 1974 - April 1975

Validation will begin as soon as the first class complete FSP. Validation will continue into the jet pilot training program (Undergraduate Pilot Training). It takes one year from SMSO graduation for the last class to graduate from UPT. Therefore, it is anticipated that full validation will not be completed until 1976.

The testing content of each phase are shown in Figure 2.

Figure 2.

PHASE I

Group Tests

Altitude Chamber

PHASE II

Group Tests (including Phase I tests)

Altitude Chamber

Individual tests

PHASE III

Group tests (including Phase I and Phase II tests)

Altitude Chamber

Individual tests

Selection Board

It is intended that tests in any previous phase would be administered in all subsequent phases. However, removal, substitution or recommended addition of tests will be a function of the validation program. The unique feature of Phase II is the introduction of individually administered testing. Convening of a Selection Board when the student reports for training starts in Phase III.

Figure 3.

PHASE I

Air Force Officer Qualifying Test (AFOQT)

Reference Test Battery

Navy Aviation Selection Board

Strong Vocational Interest Blanks

Altitude Chamber

Figure 3. (con't)

PHASE II

Phase I test

General Aviation Trainer (GAT-1)

Complex Coordination Test

Attitude Survey

Interview

PHASE III

Phase I and Phase II tests

Aviation Indoctrination

Personal Adjustment Tests

Selection Board

The specific tests in each phase are listed in Figure 3. These instruments were selected as a result of previous research and demonstrated potential. Whenever possible, a fully developed test would be identified.

AFOOT:

The AFOOT is used in officer and aircrew selection. It is administered during the recruiting process. There are five test booklets containing thirteen subtests which yield five scores: Pilot, Navigator-Technical, Officer Quality, Verbal and Quantitative. This test was validated on 1963-1965 classes. A more recent partial validation was accomplished with a sample of 1969 classes.

REFERENCE TEST:

The Reference Test Battery is an experimental battery developed by the Air Force Human Resources Laboratory. There are twenty-one brief subtests in four booklets. It is designed to yield measures of several verbal and nonverbal aptitude factors including psychomotor. These instruments will be useful in interpreting scores on the other tests.

NAVY TEST:

The U.S. Navy and Marine Corps Aviation Selection Battery is used for selection of pilots and flight officers in the Navy and Marine Corps. It consists of academic, mechanical, spatial and biographical booklets.

INTEREST TEST:

The Strong Vocational Interest Blank yields thirty-seven interest assessments. It is being used in research on career intent of junior officers in the Air Force. Previous Navy research with this instrument has shown promising results.

ALTITUDE CHAMBER:

The Altitude Chamber medical evaluation takes place before the third training week. No action is taken on any chamber problems unless these reappear later. The altitude chamber ride is scheduled as far as possible before the FSP in order to prevent any physiological difficulties from interferring with or delaying flying in the FSP.

SIMULATOR:

The General Aviation Trainer (GAT-1) is an "off-the-shelf" light aircraft simulator. However, development of the test is under contract. The purpose of the contract is to secure design and full development of a battery of "learning sample" measures which would serve as training success predictors. This contract includes design and development of the equipment system needed for administration of the battery as well as development and documentation of associated software (e.g. computer programs and syllabus) which provides the training context in which the measurement battery is embedded. Emphasis is on the problem of flying deficiency elimination (or failure due to inability to develop appropriate flight behaviors) in UPT. It is felt that an objective sample of actual learning behavior in the context of pilot training using a light aircraft simulator should yield effective predictors of training performance in the aircraft. Since the test is now under development, it will not be available until Phase II.

PSYCHOMOTOR TEST:

The Complex Coordination Test is the outgrowth of previous research to develop a new psychomotor test. Although still in use for pilot selection by several foreign countries, psychomotor testing has not been included in Air Force pilot selection test batteries since the early 1950s. Use of psychomotor tests was discontinued by the Air Force shortly after World War II. It was too costly and difficult to maintain and calibrate the many separate pieces of equipment under decentralized testing conditions. A preliminary investigation with the newly developed psychomotor test indicated that the psychomotor test is still a predictor of success in pilot training.

ATTITUDE SURVEY:

The Attitude Survey is an inventory of personality and background characteristics relevant to officer career intent. This questionnaire is the outgrowth of previous work in the area of career intent.

INTERVIEW:

An interview with a psychiatrist to assess the candidates adaptability for military aeronautics (AMA) is supplemented by interview data (biographical information) obtained by questionnaire. Previous research has shown that interview data obtained by questionnaire could be used to identify pilot training dropouts (self-initiated eliminations).

AVIATION INDOCTRINATION:

An estimated three hour orientation period will be used to give pilot candidates realistic information on pilot training and the pilot career in the Air Force. This orientation is now under development. It is anticipated that a multimedia briefing will be prepared to let candidates know what it is like to be an operational pilot in the Air Force. Assessments will be made

in conjunction with this briefing.

PERSONAL ADJUSTMENT:

These tests will address the affective domain. Emphasis will be on identifying the fear of flying, manifestation of apprehension and motivation eliminations.

SELECTION BOARD:

A Selection Board will review the candidate's records and make a decision as to whether or not he is a good risk as a pilot candidate. In addition, the candidate may be required to meet with the Board.

Validation

This is a longitudinal study involving predictive validity. Testing is done over a two day period. There will be approximately one hundred scores and subscores on the various selection instruments. Tests will be combined to compute the best total prediction. Prediction will be made for attrition and the attrition categories. Predictions will be referenced to the pilot training aircraft phases. Consequently validation for each class cannot be completed until graduation from UPT. On the other hand validation begins as soon as the class completes the light plane Flight Screening Program. Validation of each screening instrument continues until the validity has been established.

Pilot screening effectiveness will be determined and screening alternatives evaluated. Based on the results of the validation program, recommendations will be made to change pilot screening methods and selection criteria used in the operational system for selecting students to enter the pilot training program. The goal is to develop a new screening and selection system that will contribute to bringing jet pilot training attrition down to ten percent or lower.

References

Mission Analysis on Future Undergraduate Pilot Training: 1975 through 1990.
Volume 1, "Final Report Report," AFSC-TR-72-001 Vol 1, Mission Analysis Study Group, January 1972.

Mission Analysis on Future Undergraduate Pilot Training: 1975 through 1990.
Volume 2, "Introduction, Current UPT System, Training Requirements Analysis," External Influences on Future UPT. AFSC-TR-72-001, Vol 2, Mission Analysis Study Group, January 1972.

Mission Analysis on Future Undergraduate Pilot Training: 1975 through 1990.
Volume 3, "Future UPT Instructional Concepts, Future UPT Training Media." AFSC-TR-72-001 Vol 3.

Mission Analysis on Future Undergraduate Pilot Training: 1975 through 1990.
Volume 4, "Development and Refinement of Analytical Tools, Definition of Alternative Future UPT System." AFSC-TR-72-001 Vol 4.

Mission Analysis on Future Undergraduate Pilot Training: 1975 through 1990.
Volume 5, "Evaluation of Alternative Future UPT Systems, UPT Mission Analysis Findings." AFSC-TR-72-001 Vol 5.

Plan for Centralized Selection of Students for Undergraduate Pilot Training. Headquarters Air Training Command.

Programming Plan for Centralized Selection of Students for Undergraduate Pilot Training. Serial Number 2-73. School of Military Sciences, Officer. 15 March 1973.

U.S. ARMY PILOT ASSESSMENT CENTER PROJECT

Colonel Wallace F. Veaudry
United States Army Infantry School

To clarify a point at the outset, despite the title, this is not a project to assess pilots. It is a pilot project to see if the Assessment Center concept has applicability to the Army's needs.

This paper will be presented in the following order: (Slide 1)

- *Some historical uses of the Assessment Center concept.*
- *Background of events leading up to the establishment of the Army Assessment Center.*
- *The Assessment Center; mission, organization, training procedures, development of materials, exercises and scales. Pretest techniques, selection of personnel, schedule of tests and typical sequence of events.*

Before addressing the first topic, let me present a definition of an assessment process. (Slide 2)

> *"Procedure in which individuals participate in various standardized group and individual exercises designed to elicit specific skills or behaviors. These skills or behaviors are observed and scored by trained assessors".*

For example, if it is to be used to measure potential for greater responsibility or advancement to a next higher level, those skills required for future assignments will be measured.

The Assessment Center concept is not new, although industry has recently shown a great interest in this technique, as one of several, to better manage personnel. Some reports indicate that the German Army used assessment techniques as early as 1911. During WWII, the Office of Strategic Services (OSS) used this technique as part of their screening and selection process. The British Army uses the Assessment process as the criterion for entry into its officer producing military schools. The British Civil Service Selection Board was the first civilian government agency to use the assessment technique and it still uses the procedure today. In the United States, American Telephone and Telegraph (AT&T) Company was the forerunner in the use of an Assessment Center in industry. Today, as many of you know, AT&T is a leader in the use of Assessment Centers. Some other major business organizations using the technique are IBM, J. C. Penney, Standard Oil, General Electric, Ford and Kodak. US Government Agencies currently using Assessment Centers or similar techniques are US Forestry Service, Internal Revenue Service and the Federal Aviation Administration. The US Post Office uses an Assessment Center program developed by the American Management

Association. In addition a number of State and City Government agencies are using the process as are some school systems.

Now let us take a brief look at some events that led to the establishment of the project. I should not have to review for this audience the social turbulence of the 60s and its impact on our major institutions, colleges, churches and even big business. The rapid growth of technology also had its impact on the Army. Mechanized personnel records and pay systems, systems analysis and OR approaches to problem solving, the search for automated battle fields and new weapons systems all were great strides in efficiency, but at the cost of less attention to our "people" needs. An Army War College study in 1969 found that the individual at all grades perceived the leadership climate to be unfair. By direction of the then Chief of Staff, General Westmoreland, a series of actions was taken to provide for mission accomplishment with maximum concern for the individual as opposed to mission accomplishment at any price.

The US Army War College study---Leadership for the 70s---recommended as one of its solution concepts to the problems it identified, use of contemporary scientific approaches to leadership instruction. The CONARC Leadership Board in its report "Leadership for Professionals" also stressed the need for updating leadership instruction in service schools through greater use of recent advances in the Behavioral Sciences. As the Vietnam War was coming to an end and an all Volunteer Army started to emerge, the impact on the Army had to be recognized. A smaller Army had to be more efficient and more effective. A Volunteer Army had to place greater emphasis on leadership at all levels if it was to be successful. The unending flow of "new blood" as needed, no longer existed. Quality became the 'by-word' as quantity rapidly diminished.

By direction of the Army Chief of Staff, an Assessment Center was established at Fort Benning, Georgia. Initially the Center came under the guidance and direction of the Office of the Special Assistant for Training (OSAT) in the Office of the Chief of Staff while assigned to the US Army Infantry School. At present guidance is received from the Office of the Deputy Chief of Staff for Personnel through the Director of Human Resources. In November 1972 the first officer and one NCO were assigned to the Assessment Center and directed to find a building and get organized. By February the majority of the staff had reported and in March the present director arrived to manage the Center.

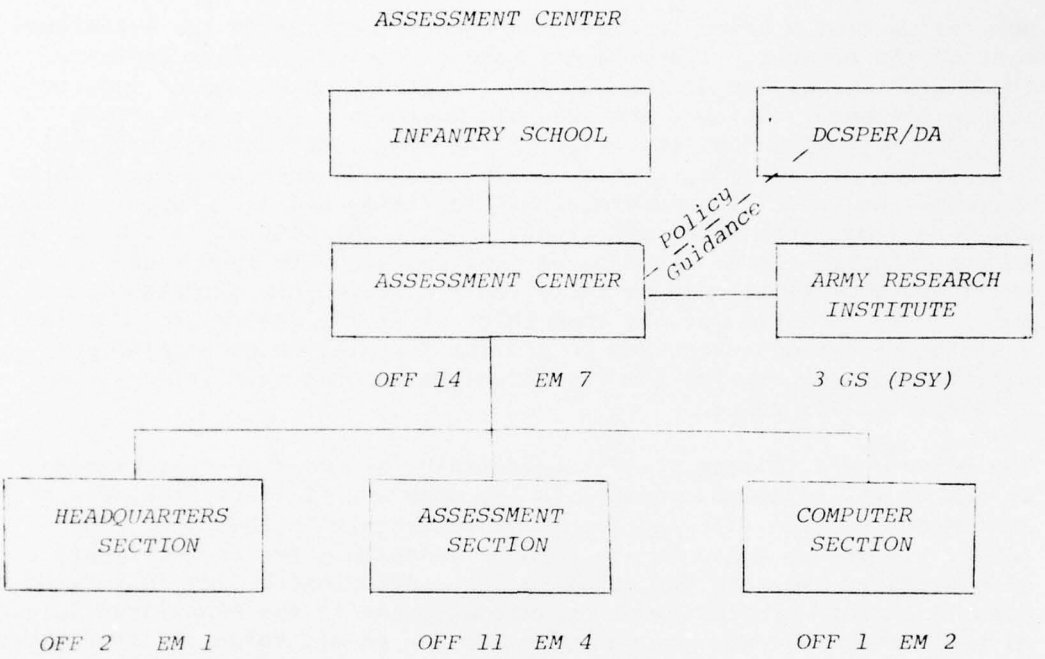
The mission of the Center is as follows: (Slide 3)

a. *Test the assessment concept as a precommission screening device.*

b. *Test the assessment concept as a leadership and career development technique.*

c. *Determine the feasibility and applicability for Army wide use of the concept and techniques developed under a and b.*

In order to accomplish the mission, the Assessment Center is organized as follows: (Slide 4)



The director reports to the Assistant Commandant of the Infantry School. ODCSPER/DA provides guidance and is the Department of the Army point of contact. The Army Research Institute provides technical advice and support through a three-man field team of professional research psychologists co-located with the Center. The remainder of the organization, Center headquarters, assessment section and computer section are self-explanatory. Computer hardware support is provided by the Management Information Systems Office (a computer center) at Fort Benning.

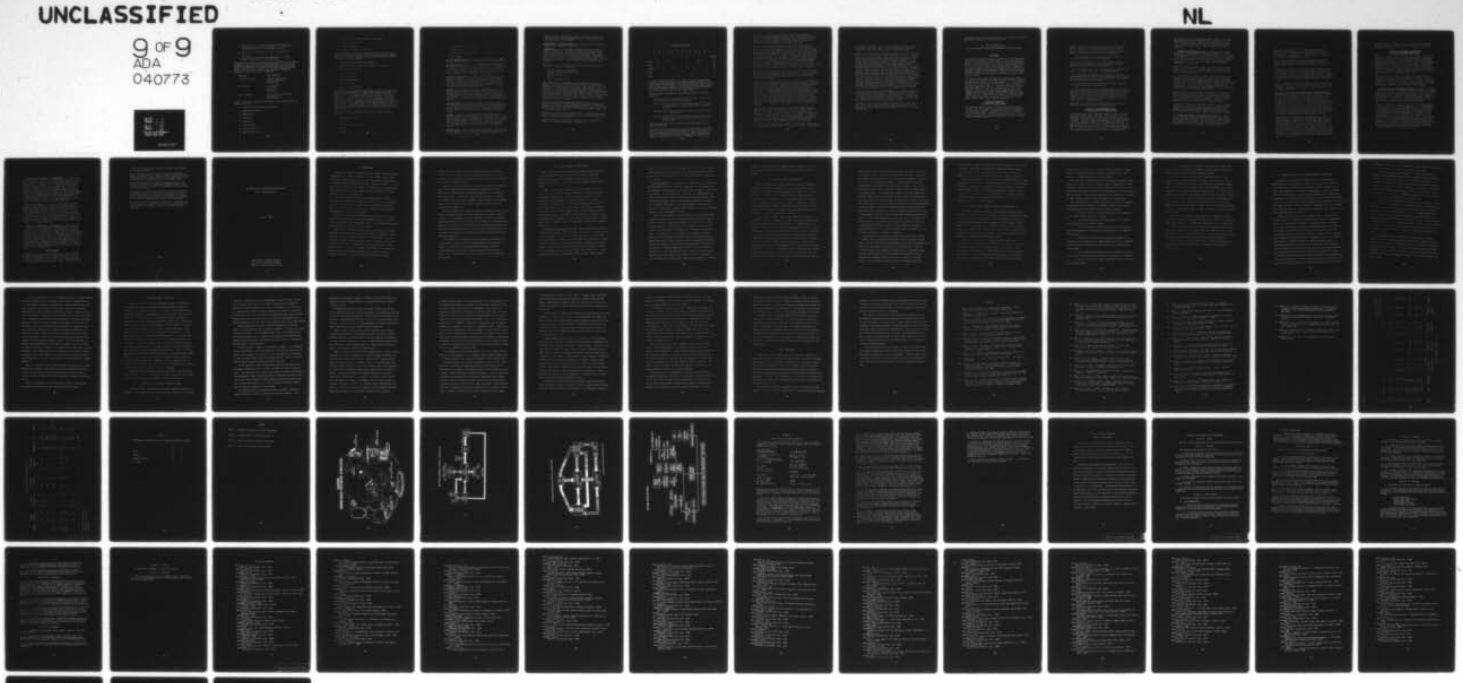
The total staff of the Center consists of fourteen officers, seven enlisted men and three professional civilian psychologists and three secretaries from Army Research Institute (ARI). The majority of military personnel (eleven officersand four NCOs) are assigned to the Assessment Section as assessors with nine of these officers and two NCOs also acting as counselors. With one exception, all of the officers have master's degrees in the social or behavioral sciences and one officer has a Ph.D. in psychology. These officers were selected by name through OSAT/DA. The noncommissioned officers were selected based on experience and special abilities or qualifications. They are an exceptional group.

The time schedule established for the assessment project is shown on this next slide. (Slide 5)

AD-A040 773 AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TEX F/G 5/9
PROCEEDINGS, ANNUAL CONFERENCE OF THE MILITARY TESTING ASSOCIAT--ETC(U)
NOV 73

UNCLASSIFIED NL

9 OF 9
ADA
040773



END

DATE
FILMED
7-77

a. *Feb 73-30 Jun 73 - organization, training of personnel, development and pretest of techniques, tests, situations and measurement scales, establishment of program.*

b. *Jul 73-Jun 74 - assessment and test of techniques and devices.*

c. *Jul 74-Jun 76 - follow-on study to determine validity of concept, situations, tests and techniques.*

One of the first steps in developing the test materials was to determine the behaviors that were to be elicited and measured. A review of the literature including the research at Ohio University and the Army War College resulted in a list of dimensions for leadership and establishment of a series of gross indicators for these major dimensions. Examples are shown on this slide. (Slide 6)

Dimension	*Gross Indicator*
Communication Skills	*informal and formal oral* *written communications* *understanding*
Forcefulness	*self-confidence* *energy level* *risk taking* *initiative*
Decision Making	*use of available information* *decision quality* *decisiveness*

A total of 12 broad categories or dimensions are used, as shown on this slide. (Slide 7)

1. *Social Skills (Interpersonal Competence)*
2. *Communications Skills*
3. *Adaptability*
4. *Motivation*
5. *Forcefulness*
6. *Mental Ability*
7. *Decision Making*
8. *Administrative Skills*

9. *Effectiveness in Organizational Leadership Role*

10. *Supervisory Skills*

11. *Physical Fitness*

12. *Technical and Tactical Competence*

As the exercises were developed and tested, specific indicators were designed and scales developed for each of the situations. For example, under effectiveness in interpersonal situations the following specific indicators are used: (Slide 8)

Effectiveness in Interpersonal Situations

Demonstrates skill in handling interpersonal situations

Interacts with individuals

Interacts with group

Assesses group norms

Actively participates

Accepted by others

Avoids arousing hostility

As each exercise or situation was developed, a series of test runs was conducted using personnel of the appropriate grade and experience from those assigned to the Infantry School, both students and faculty. Adjustments in instructions, complexity, and measurement scales were made until we were satisfied that we had a challenging problem that elicited the behaviors we wanted and that these behaviors could be appropriately scored. Outside assistance was obtained for some problems or situations. The HumRRO element at Fort Benning developed three simulates which I'll discuss later. The Conglomerate Game was purchased from Byham-Bray Associates, and selected standard psychological tests were purchased.

The types of exercises used by the Center are shown on this slide. (Slide 9)

Paper and pencil tests

Interviews

Leaderless group discussion

War game

In-Basket

Combat or emergency simulate

Writing exercise

Conglomerate (management exercise)

Field exercise

The Paper and Pencil Tests are a battery of psychological tests which measure intellect, leadership, vocational interest and reading ability.

The Interview (two types) - one is a background interview to determine motivation, experience, problem areas, and the assessee's knowledge of his strengths and weaknesses. The second, an Appraisal Interview, is a training exercise where each assessee interviews two others to select one for a position within a battalion. This interview measures organization of thought, communication skills and social interaction.

Leaderless Group Discussion - consists of different problems for each category of assessees. For Senior NCOs it is a combined individual and group task where six assessees are assigned a mission of getting a soldier from their unit selected as Soldier-of-the-Month and the group being assigned a mission of selection of one soldier from the six. This exercise measures leadership, forcefulness, persuasiveness, organizational ability and group interaction.

War Game - is for officers only. This is an assigned role rotating leader exercise conducted in two 2-hour increments. Teams of six players engage in cost effective analysis in a military force planning environment. Total costs, R&D, intelligence acquisition, balanced offensive/defensive forces are all considered under limited budget and time constraints. This exercise measures organizational and leadership behavior.

In-Basket Exercise - a typical 1SG, company or battalion commander in-box is presented to the assessee who has 3 hours to address each item in the in-box. This exercise measures problem solving, decision making, work organization and leadership. It is followed by an interview to discuss reasons for action taken and the relationship of some of the actions.

Simulate - consists of a different exercise for each category of assessee. NCOs act as platoon leaders of a combat patrol with a mission change requiring an air mobile operation. It is a 5-hour exercise using radios as the only means of communication. It measures stress tolerance, risk taking and leadership. Senior company grade officers play the role of Company Commanders in a similar exercise. The officer candidates and junior officers play similar roles in a civil emergency setting.

Writing Exercise. There are two forms of this exercise. The first form requires the assessee to write a statement recommending discharge of one of his subordinates. The second form is a staff action paper. Both

exercises measure the assessee's ability to marshall necessary facts from an ambigious situation and relate them in a concise, complete and grammatically correct fashion.

Conglomerate - a management game used to measure emergent leadership, aggressiveness and social interaction.

Field Exercise - a 5-hour, leader designated, exercise involving a team of six assessees. There are six lanes with a mission and an obstacle in each lane. A choice of materials is provided. It measures leadership, planning, organization and stamina. Through the design of the exercises and by means of questionnaires, interviews and peer ratings, a tri-focal view is obtained on each individual. His own evaluation of himself--his peer ratings--and the evaluation of the assessors.

A poem that I saw in a magazine not long ago seems to sum up what we are after. I do not know the author.

> And there were three men went down
> The road as down the road went he.
> The man they saw.....
> The man he was .....
> and
> The man he wanted to be.

Concurrently with the development of the exercises, training was conducted for the assessor staff. Outside assistance was obtained from a civilian firm (Development Dimensions, Inc.) that conducted a three-day training session at the Center. In addition selected professional counselors, to include Baker Consulting Associates, presented instruction to the entire staff. In-house training consisted of each member assisting in the development of the exercises, taking the tests or participating in the exercises and studying the scales. Then a series of training exercises was held using local personnel as assessees with the center staff conducting the assessment.

Since we are assessing five different categories of personnel, the exercises are prepared with different levels of complexity. Problems are designed to "stretch the minds and abilities" and are cast in areas that the individuals should encounter in their future assignments rather than in their present experience.

Assessees are divided into three groups of six for team exercises and to provide flexibility in conducting the exercises within the capabilities of the thirteen-man assessor staff.

This slide shows the categories of individuals to be assessed and a schedule of assessment for the year. (Slide 10)

ASSESSMENT POPULATION

	J U L	A U G	S E P	O C T	N O V	D E C	J A N	F E B	M A R	A P R	M A Y	J U N	TOTAL
IOAC		X		XX							XX		108
IOBC		X	X		X		X	X			X		108
ANCOES	X		X		X			XX		X			108
BIOCC			X				X			XX		XX	108
ROTC						XX			XX				72

A total of eighteen individuals are randomly selected from Department of the Army lists of personnel assigned to attend the several courses at the US Army Infantry School. The orders for these personnel are annotated to indicate participation in the Assessment Center pilot program for a four-day period. They are ordered to report to Fort Benning one week in advance of the start of their regular course.

A typical sequence of events at the Assessment Center is as follows: (Slide 11)

Three days prior to assessment - assessees report to Fort Benning for in-processing at their student company.

1st day - assessees report to the Assessment Center at 1500 hours for in-processing orientation and to begin assessment exercises.

2nd and 3rd day - assessees undergo assessment from 0700 through 2100 hours.

4th day - 0700-1700 hours, assessees complete exercises at 1700. They fill out an exit questionnaire and are released at 1730 hours.

(OCS candidates begin their assessment days at 0530 with an hour of physical training.)

The facilities of the Assessment Center, other than office space, consist of ten small air-conditioned assessment rooms, large enough for a six-man assessee team with three observers, a large conference room where all eighteen assessees can be assembled, and a video monitor cubicle containing six TV monitors. In addition, a field area with

six lanes, each containing a different obstacle, was constructed. There is billeting for eighteen assessees and a small lounge for relaxation as time permits. We have the capability of video taping in all assessment rooms with remote TV cameras and video cassette recorders. These tapes prove invaluable in the counseling sessions.

For those personnel being assessed for Leadership Development, a detailed, one-on-one counseling session is provided within a week to ten days following the three-day assessment cycle. The Infantry School class and our assessment schedules dictate availability of both assessees and counselors for scheduling of these sessions. The rating scales are all designed for keypunch and a computerized profile is provided to the counselor on each man he is to counsel. In addition, qualitative data is provided in reports from the assessors and from the entry interview. The counselor then prepares a synopsis sheet on his assessee covering relative strengths and weaknesses in the areas measured. By careful editing and clipping of the video tapes, examples are selected from the actual exercises to assist in the counseling sessions. These sessions are open ended, but usually last from 3 to 4 hours to each assessee.

What has been our experience to date? As seen on Slide 10, "the assessment population", we have processed three groups of captains, two groups of lieutenants, two Senior NCO groups and one officer candidate group through the assessment cycle. We won't get our first look at ROTC students until December.

It is too early to talk in definite terms of results, but I can speak of trends and early indications concerning the Assessment Center as a leadership development technique. You recall that Senior NCOs and captains enroute to the Career Course, and lieutenants enroute to the Basic Infantry Officer Course all received counseling feedback based on the assessment. In addition, each assessee completed a questionnaire at the end of the assessment and upon completion of his feedback session. Without exception, every NCO and officer found the experience to be both challenging and rewarding. But more importantly, they all commented positively about the training value of some of the exercises. All believed that they had a better understanding of themselves and in almost all cases readily accepted the evaluation of strengths and weaknesses presented by our counselors. A few disagreed on minor portions of our findings.

Let me digress a moment and discuss the counseling feedback technique. Ideally, the counselor elicited from the individual what he learned about himself during the assessment cycle--what each felt his own strengths or weaknesses were. The counselor then reenforced those that our findings agreed with and discussed in greater detail any areas not volunteered by the assessee. As I mentioned previously, video tape is used to demonstrate specific areas of strengths or weaknesses when appropriate behaviors were recorded during the actual exercises.

As for prediction of success as an officer candidate or for commissioning, with only one group tested it's too early to comment. We are getting

the behaviors we hoped to elicit. It is easy to discriminate among individuals. Our scales appear to provide good measures, but we do not have sufficient data to go beyond these general statements.

However, by the end of the test period we expect to be able to recommend the best use of the Assessment Center technique for the Army. We will have sufficient data to address specifics in terms of staffing, organization, types of exercises that lend themselves to general use and essential qualifications or training of assessor personnel. As part of our experimental design, we hope to be able to run serveral groups of individuals without behavioral science backgrounds or degrees through the Center as assessors. With minimal training on the exercises and the scales as well as some hints on observing behavior, we hope to be able to prove it doesn't take a master's degree to become an effective assessor. We will make this test by comparing results between our trained assessors and those individuals we use on an experimental basis. One rather obvious fall-out from the experience to date is the invaluable training the assessors are receiving from their participation in the program. There is no doubt in my mind that these officers and NCOs will find that the experience of close observation of behaviors, detailed study of individuals strengths and weaknesses and the practice of counseling techniques will not only enhance their ability in any future assignment, but make them better individuals, too. You can't study other people without doing a little introspection and self-analysis.

Finally, there will be a two-year follow-up of our results from July 74 to July 76. This research will be conducted by the Army Research Institute Field Unit at Fort Benning. Through analysis of the data we will have generated on individuals and groups as well as through study of records, questionnaires, peer ratings and possible interviews, ARI will determine if our results and findings are valid.

Gentlemen, I could go on for another hour about this exciting project, but I know there are many other topics equally as exciting to their sponsors. I will be glad to answer any questions at this time or later during breaks.

Assessment Centers: Validation of their use for selection, training, and career development in a military setting.

KAY H. SMITH, Ph.D.
Chief, Fort Benning Field Unit
U.S. Army Research Institute for Behavioral & Social Sciences

BACKGROUND

In late 1972 an Assessment Center was established at the directorate level within the US Army Infantry School (USAIS) at Fort Benning, Georgia. Initial planning for the Center was a joint effort of the Office of the Special Assistant for Training (OSAT) under the Army Chief of Staff, and the US Army Research Institute for Behavioral and Social Sciences (ARI), a class II activity under the Office of the Chief of Research and Development (OCRD). At the time that the Assessment Center was established at the USAIS, ARI set up a field unit at Fort Benning. R&D support for the Center was its initial primary mission.

The details of development and operation of the Assessment Center are more completely given in the paper prepared for this conference by COL Wallace F. Veaudry, the Center Director. I will confine my comments on the Center itself to conceptual issues. The Center was tasked with two missions: 1) Prediction of success in Officer Candidate School (OCS), and later success as an officer, and 2) Career development, through assessment followed by performance feedback counseling, for Infantry Officer Advanced Course (IOAC) students and for Advanced Noncommissioned Officer Education System (ANCOES) students. During Center development the task was broadened to include assessment of a sample of Infantry Officer Basic Course (IOBC) students and a sample of ROTC cadets.

ASSUMPTIONS UNDERLYING THE ASSESSMENT CENTER CONCEPT

The Assessment Center concept is not particularly revolutionary, nor is it new. The series of performance requirements placed upon candidates for knighthood in medieval times might be categorized as an Assessment Center. In modern times the Office of Strategic Services' World War II selection process described by Murray, et.al. in The Assessment of Men stands as the prototype for many current

efforts. In recent years considerable use of Assessment Centers has been made by members of the industrial community, notably AT&T, J.C. Penney Co., IBM and General Electric. A number of basic assumptions seem to underly the use of Assessment Centers. Among the more important assumptions are the following:

1. The characteristics to be measured can be identified, elicited at the Center, and reliably observed.

2. The characteristics to be measured require some sort of performance or situational testing, and that easily portable, paper and pencil tests alone cannot do the job.

3. The characteristics to be measured require interaction with people, either other examinees or role-playing assessors.

4. Situational "job samples" will provide indications of examinee work behaviors which are more standardized and controlled than those shown in his real-life job, and thus will make comparisons between examinees more meaningful.

5. Observations by trained assessors will be more reliable and valid than those made by supervisors.

It is entirely possible that in a given application of the assessment center concept some or all of these assumptions may be invalid. Since centers are usually required to measure such elusive characteristics as communication skills, personal impact, sensitivity, leadership, etc., it may be difficult to meet assumption number 1. If this assumption is not met, of course, there can be no justification for use of Assessment Centers.

Measurement of reliability and validity become then essentially the process of proving that the assumptions are true.

VALIDITY OF THE ASSESSMENT CENTER AS A PREDICTOR FOR SELECTION PURPOSES

As mentioned above, one of the Center tasks is to provide measurements which could serve as selection criteria for OCS. The basic problem here seems relatively simple, since the criterion to be predicted---successful completion of OCS---is concrete and readily available. The task is complicated somewhat by the hope expressed by some high-ranking Army officers that the Assessment Center would allow the Army to weed out men who now successfully complete OCS but later show flaws as officers. Without going into the problems of measuring such things as "character", these hopes suggest that successful performance in OCS is not to be regarded as an

ultimate criterion. The validation problem, so far as OCS cadets are concerned, seems to be threefold: 1) Prediction of level of performance in OCS, 2) Prediction of quality of performance as a commissioned officer, 3) Evaluation of OCS training procedures and performance criteria. Let's consider each of these separately.

Prediction of performance in OCS. Currently success in OCS is dependent on performance in three general areas: 1) traditional paper and pencil tests covering academic areas, 2) ratings of leadership skills by peers and tactical officers, 3) persistence in an intensive and highly demanding task requiring total involvement over several weeks' time. The latter is reflected in the measurements listed as 1 and 2 but also is shown by voluntary withdrawal from the program.

While the Assessment Center utilizes a battery of standard paper and pencil tests which will probably predict performance in academic areas quite well, it is obvious that such testing does not require an Assessment Center.

It is in the areas of leadership skills and task persistence that the Center shows the greatest promise for accounting for unique variance. The situational tasks of the Center include opportunities for observation of emergent leadership as well as for evaluation of performance in assigned leadership tasks. In each of these situations multiple ratings are made by trained assessors and peer evaluations are also obtained. These should provide good prediction of the similar ratings obtained later in OCS.

Since the Center is a very demanding and totally involving experience covering a three and one-half day period, it also should provide an opportunity for observing task persistence and dedication under carefully standardized conditions. Coming, as it does, in close proximity to the beginning of OCS it should provide a unique opportunity for prediction of this elusive factor.

The problem of predicting performance in OCS is a fairly straight forward one. A major difficulty which must be faced is that our limited sample size will probably preclude empirical predictor weighting with cross validation. Hopefully, the Center will be maintained beyond FY74 in order to obtain further validation data.

Prediction of quality of performance as an officer. Validation of Assessment Center scores against criteria of successful officer performance involves all the problems that these criteria have demonstrated in the past. Officer Evaluation Reports (OERs) are inflated, are not executed consistently by different raters, and show very little discrimination. Used alone, OERS are

inadequate as indicators of officer potential. Graduates of OCS go to widely varying assignments, both by branch of service and assignments within branches. (The Army OCS is currently branch immaterial. Graduates go on to the officer basic course in the branch to which they are assigned.)

Data will be available on performance in branch officer basic courses; however, the curricula and evaluation procedures vary and the numbers assigned to individual branches will be quite small.

The major source of data on officer performance will have to come from specially designed peer, superior, and possibly subordinate ratings made after a specified period of time in the man's first officer duty assignment. The variability contributed by difference in assignments may be partially accounted for, by using procedures for grouping of junior officer assignments previously developed by ARI.

One of the problems to be solved will be the degree to which the ratings can be made to reflect performance on specific dimensions of behavior. Research on this problem is currently underway in a joint ARI-HumRRO effort.

Evaluation of OCS Training Procedures and Performance Criteria. The third area of use of Assessment Center data follows naturally from the two just discussed. Since we will have three sources of performance evaluation data, it is logical to compare them. The question being asked is this. When Assessment Center data and OCS performance scores made different predictions about performance of an officer assignment , which has the greater probability of being correct? This, of course, assumes that there will be discrepancies. The more basic question is this. Are the characteristics required for successful performance at OCS the same as those required for successful performance in an officer assignment? To the extent that they are not, the next question becomes; Can we account for the differences with Assessment Center data?

To some extent, it is hoped that some light can be cast on these areas of concern. Small sample size will again make definitive answers difficult to achieve. An area of real concern involves attrition from OCS. This further reduces the sample insofar as the officer assignment performance criterion is concerned. It will provide positive validation data in the case of those with poor Assessment Center scores. The question with regard to those who do not complete OCS, but who have good Assessment Center scores becomes moot. If they fail in OCS they become negative validation cases, but if they voluntarily withdraw the questions about how they might have performed in OCS and later in officer assignments

remain unanswered. Perhaps at least we may find some data which will shed light on why they withdraw. Sample size will again be a severe restriction.

VALIDITY OF ASSESSMENT AND COUNSELING PROCEDURES FOR CAREER DEVELOPMENT

While the problems discussed in the previous section are difficult they are at least familiar. The criterion problem is always with us whenever performance is to be predicted. The validation problems now to be addressed include all those problems but are further complicated by still others. It is still necessary to demonstrate that the behaviors measured at the Center relate to significant dimensions of job performance, but now we must also show that going through the assessment process and receiving feedback and counseling on his performance produces measurable effects in a man's later performance and/or his subsequent career choices.

While in the previous section the population to which we were referring involved OCS cadets, this section will have as its main focus NCOs from ANCOES classes and captains from IOAC. To a lesser extent the discussion will also apply to 2nd lieutenants from IOBC. The scope of this paper will not permit detailed discussion of the problems and procedures unique to each of these population.

Validation of the measurement dimensions. Procedures for demonstrating the validity of the Center measurement process will be much the same for the counseled groups as for the OCS cadets. Assessment Center scores will be compared with performance in the courses at USAIS, and later with ratings of performance in duty assignments. This process will take some time with the IOAC captains since they are assessed at the beginning of a 39-week career course. The other courses are much shorter, ranging around 13 weeks.

The major complication with this validation of the basic measurement scales involves the feedback counseling. It is assumed that, in some cases at least, learning about weaknesses will lead to activity designed to strengthen that area. To the extent that this is true correlations between Assessment Center scores and later criteria should be reduced. Unfortunately, constraints on time and availability of subjects did not allow a design involving an assessed but not counseled subsample within the same population as an assessed and counseled sample. We may be able to get a partial answer to this by comparing validity coefficients of OCS cadets and IOBC 2nd lieutenants for officer assignment performance scores. This will be tenuous at best since there are several significant differences between these groups which could account for any observed differences in the validity coefficients.

Effects of Counseling on later performance. Since we do not have an assessed but not counseled group what we will be measuring here is the effect of the total assessment and counseling experience. The first question will be; Do counseled students do better at USAIS than uncounseled students? Since the Assessment Center sample is a random selection from the USAIS class, it should be possible to compare their school performance scores with the remainder of the class. The kinds of school performance measures available vary from population to population. The IOAC and ANCOES scores focus on academic type tests while the IOBC classes produce in addition to academics some relatively good leadership scores from peer and tactical officer ratings. The constraints of the school setting are such that it is unlikely that we will be able to obtain specially constructed performance measures.

As with the OCS students, ratings will be obtained on their performance in post-school duty assignments. For these purposes a carefully selected control group will be chosen from each class where students were assessed. These people will be followed up in the same fashion as those assessed. In the case of the duty performance ratings it will be possible to design the measures in such a way as to focus on particular weaknesses which were noted such as reading and writing skills among the NCOs. These will have been a focus in counseling sessions and the performance ratings should provide data on whether the assessees actually did anything to attempt to correct deficiencies.

Effects of counseling on career decisions. This will be the most difficult area in which to get anything approaching definitive answers. It may be possible to get information concerning requested assignments and then to compare these with Assessment Center profiles. It may also be possible to get comparison figures for control and counseled groups on such things as leaving the Army. This will be particularly interesting for IOBC students, but will require waiting 3 years in most cases. The greatest bulk of data in this area will probably have to come from questionnaires covering career plans. One purpose of such questionnaires would be to assess the extent to which there is a residual of the profound feelings of self-understanding and new insight typically expressed at the time of the counseling.

OTHER RESEARCH

The above discussion outlines in very general terms some of the definite research plans which have been developed for validation of the USAIS Assessment Center. Some of these are not cast in concrete at present and in a moment I will be very happy to hear comments and new ideas from the floor. Before I do this however,

I would like to mention briefly some of the other research questions we are addressing.

On the assessment process we will be looking at the following areas: 1) Development of a standard training program for assessors, 2) Adaptation of assessment situations for training pruposes, and 3) Determining the dimensionality of our measurement space.

On the counseling process our research goals are two-fold: 1) Experimental comparison of alternative feedback methods in order to discover the optimal method, 2) Development of a standard counselor training program.

Finally we will be looking at profiles of scores for our different assessment populations. These will provide valuable input to USAIS regarding areas which should receive emphasis in training courses. We also are accumulating very interesting normative data on the commercial paper and pencil tests in use at the Center.

Now we get to the main purpose of my being here. Please feel free to make any comments, suggestions or criticisms regarding the research problem of validation. Be as creative and/or fanciful as you like.

A PROPOSED SPATIAL ORIENTATION/DISORIENTATION FLIGHT TRAINING CONCEPT

Patrick J. Dowd

Environmental Sciences Division
USAF School of Aerospace Medicine
Aerospace Medical Division (AFSC)
Brooks Air Force Base, Texas 78235

I. INTRODUCTION

Within man, a highly sophisticated inertial guidance system has evolved phylogenetically. However, his system is rather limited in that it does not provide enough information, and at times its vestibular organs elicit convincing erroneous information, when man is flying an aircraft. It is the duty of flight instructors and trainers of student pilots to compensate for these deficiencies within man so they can avoid the sometimes fatal consequences of spatial disorientation.

There is increasing agreement that the present spatial orientation/disorientation training is not meeting the needs of the pilots. Therefore, those of us interested in the complex area of flight safety and training must not allow ourselves to be bound by what exists; we must make our own recommendations in terms of what should be.

Historically, spatial disorientation has been given little or no consideration in training plans--in affecting adequate exposure to illusions or other activities aimed at informing the pilot of the relationship spatial disorientation has to flight, to his person, etc.

In his natural environment man, by using his functioning sense organs and muscular effectors, is an almost perfect cybernetic mechanism for maintaining his position, attitude, or motion. In flight, however, man is far from perfect in this respect, especially when he is deprived of good visual contact with the earth. The degree of his imperfection depends, in part, upon his ability to respond to instruments which present him abstracted and reduced information regarding these parameters and upon his skill in interpreting messages from his vestibular and proprioceptive systems.

Maintaining proper orientation in a high performance aircraft in weather is one of the most complex, nonverbal, intellectual functions that a man can perform. It is one that is extremely difficult to study in situ because of the speed and distance involved.

Laboratory studies and adequate on-the-ground training require abstraction of the variables involved and working with combinations of these abstractions, culminating with an ultimate integration of all variables in the aircraft itself. A great many efforts in the past have failed because of inadequate definition of these variables, undue emphasis on the vestibular system as it is known, reliance on face validity as a criterion in the absence of better data, etc.

This paper suggests how a spatial orientation training device (1, 2) can be used in conjunction with lecture and slide material to meet a threefold purpose: (1) The student will learn to anticipate the kinds of perceptual experiences he will receive in the air; (2) the conditions of and mechanisms whereby these experiences are mediated will be elucidated for him; and (3) he will practice maintaining control of the device and performing desired maneuvers under conditions which will disturb his orientation systems and lead to erroneous information.

A brief review of the causes and costs of spatial disorientation, the present status of spatial orientation indoctrination, and the research studies on the USAF School of Aerospace Medicine spatial orientation trainer (SOT) will be discussed. Also, the proposed training concept of using a SOT in flight training for spatial orientation indoctrination will be presented.

II. CAUSES OF SPATIAL DISORIENTATION

Man has inherited and developed complex sensory systems which enable him to move purposefully over the surface of the earth. However well adapted these systems are to terrestrial activity, they do not always function well in aircraft. Let us review the three main sensory systems briefly.

Vision: This is man's dominant sense. A pilot flying on a clear, sunny day and enjoying a well defined horizon has little or no difficulty with spatial disorientation. Under other conditions, however, vision can contribute to disorientation. A tilted cloud bank may be accepted as a true horizon. In the absence of a well-defined horizon, ground lights may become confused with stars, or a well-lighted shoreline may be accepted as the horizon. Pilots have been known to confuse a star or distant ground light with wing lights. The lights of other aircraft do not allow for much depth perception or judgment of rate of closure, leading to midair collisions. When a pilot is viewing a single stationary light, he might project imagined motion to it.

Proprioception and tactile sets: Man's sense organs provide information concerning the activity of skeletal muscular groups and of the relative position and movements of his limbs. The senses provide the recovery feedback for skilled movements. However, the various subsystems are subject to changed loading by acceleration; further, their activity is modulated by vestibular inputs. Tactile and deep pressure senses inform the individual of contact with other bodies, that is, the earth, seats, control sticks, etc. In aircraft, the results of such stimulation can be erroneous. Being pressed

against the seat back during an accelerated takeoff roll can give the pilot the impression that the seat and aircraft are tilted backwards; pressure against the seat and shoulder belts in a down draft can make him feel that he is inverted.

Vestibular System: The vestibule of the inner ear serves in stabilizing the head during locomotion. One set of organs--the otolith organs--senses the attitudes of the head relative to gravity and is an analog of the artificial horizon on the instrument panel. The other set of organs--the semicircular canals--senses angular accelerations imposed upon the head in any plane during walking or running. Otolith organs and semicircular canals send coordinating information to the skeletal muscles, via the brain, in such a way that the skull provides a stable platform for the eyes in walking, running, or even more vigorous activity.

When the head is turned or tilted deliberately, the vestibule assumes a different function; it sends information of this movement to the brain stem to provide for a compensatory movement of the eyes in their sockets so that fixation on the viewed object is not disrupted. For this kind of motion the semicircular canals signal rate of turn in the appropriate plane. If the motion is of short duration, the response is accurate. However, due to the mechanical features of the canal system, which integrates accelerations of short duration in order to generate the transient rate-of-turn signal, the system does not register turns of long duration. In this case, it confuses completion of a turn in one direction with the initiation of a turn in the opposite sense. Another very serious error generated by the canal system is the response to Coriolis stimulation; if an individual moves his head while riding turning equipment, his canals

generate a complex response, the dominant feature of which is a rate-of-turn signal in a plane other than that of his turn or that of his head movement.

III. COST OF SPATIAL DISORIENTATION

Spatial disorientation accidents continue to concern all military, civilian, and commercial airline pilots. A U.S. Navy report listed 395 naval aircraft related to spatial disorientation accidents over a period of 8 years, 1958 to 1966 (3). According to Smiley (1967) the primary "cause of real concern for safety in aircraft" was spatial disorientation for 224 Canadian military pilots (4). In 1971, a conference was held by the NATO AGARD Aerospace Medical Panel Meeting in Luchon, France, on "The Disorientation Incident." This AGARD report (5) highlighted the present seriousness of spatial disorientation in aviation.

Spatial disorientation is costly in lives, material, and money, as illustrated in Table I. Figures for the USAF losses in 1971 represent only those thus far identified; to some extent, data for 1958-1967 may be low because of the difficulties of accident investigations in Vietnam. Spatial disorientation is involved in a disproportionate number of fatal accidents.

Air Force data show that overall safety has improved significantly; 1955-1958 and 1958-1969 records show that the annual rate of all accidents was reduced by 72% and of fatal accidents by 82%, yet the annual rate of spatial disorientation accidents decreased by only 16% and fatal disorientation accidents by only 47%. Comparison of 1969-1971 data with that of 1958-1969 shows that the annual rate of all accidents was reduced by 10%, but there was an 84% increase for fatal accidents; however, the annual rate of

spatial disorientation accidents decreased by 50%, and of fatal disorientation accidents by 59%. The increase in complexity and power of the aircraft has probably increased the dollar cost of spatial disorientation even though the rate of loss has diminished slightly. This table shows that spatial disorientation is a problem to other air services as well. Except for two fixed-wing aircraft, the U.S. Army spatial disorientation accidents refer to rotary wing aircraft. In reviewing other studies on spatial disorientation problems, Tyler and Farr (14) reported that 96% of 2000 USN pilots reported one or more disorientation experiences, 38% were hazardous; Clark (17) reported that 93% of 336 pilots of all U.S. Services reported one or more disorientation experiences; Smiley (4) reported that 82% of 244 RCAF pilots reported experiencing vertigo or disorientation, 8.5% of 144 pilots experienced one within a 1-month period, and 6.7% of 244 pilots experienced hazardous disorientation within a 1-year period; Doppelt (12) reported that 7.5% of 55 USAF pilots reported disorientation episodes experienced within 1000 flying hours; and recently Rayman (18) listed a total of 26 USAF accidents resulting from spatial disorientation for the period of January 1966 to November 1971. Out of those 26 accidents, 16 resulted in fatalities.

Recent progress in cockpit and instrument design has greatly assisted in lowering some of the potential disorientation hazards during instrument flight. New and more easily read instruments are constantly being introduced. Reduction in the number of instruments for "cross checking" to ascertain aircraft control has also enhanced the safety factor in instrument flying. The relocation of radio and aircraft controls minimized the potentially hazardous head movements required during critical phases of instrument flights. All this progress, however, appears to be somewhat negated by

the increasing performance capabilities of today's aircraft. To see a result of technical advancement, one only needs to examine the various stresses a pilot undergoes in the ever increasing accelerations of the USAF F-15 aircraft, or even in some commercial aircraft such as the 747 on takeoff.

Spatial disorientation is the most frequently reported biomedical incident threatening aircraft crews; it is implicated in 10% of Air Force fatal aircraft accidents recently studied, 1969-1971 (6). This has been made a human-factor problem almost by default, since flight instrumentation supposedly capable of providing the pilot with sufficient information for completely safe IFR operations has proven to be prohibitively expensive in initial cost and in maintenance, and also of low reliability.

IV. PRESENT STATUS OF SPATIAL ORIENTATION INDOCTRINATION

Within the Air Force, the problem confronting flight surgeons, physiological training officers (PTO), and others interested in eliminating spatial disorientation accidents is threefold: first, to demonstrate effectively to pilots that spatial disorientation can happen to anyone; second, to discern through testing which pilots might be prone to lose aircraft control under disorienting maneuvers; and, third, to offer pilots a way to become resistant to the effects of disorienting influences. The SOT was designed to study the problem of disorientation from these three aspects.

Current Air Force practice is to expose the pilot once a year to indoctrination films and lectures, and then to fly him through a set series of maneuvers under the hood, returning control of the aircraft to him at certain points in the maneuvers. The maneuvers listed in the flight training syllabus are: climbing while turning (posturogravic- or somatogravic-type illusion), diving during turn recovery (posturogravic-type illusion),

tilting to right or left (leans), reversal of motion (similar to a spin), diving or rolling beyond vertical (Coriolis or G-excess effect), and sensation of climbing (posturogravic-type illusion).

Some bases add the demonstration of motion-induced illusions by using a Bárány chair. Either each pilot rides the chair and experiences the false sensations of turning and/or the Coriolis illusion, or one pilot is used to demonstrate his reactions and relate to the observing pilots his false sensations of turning or rolling or pitching while being rotated about the vertical axis and tilting his head in either the sagittal or lateral planes.

The undergraduate pilot training (UPT) instructor staff and students at Laredo AFB (T-37 and T-38 phases of training) reported on their spatial orientation/disorientation indoctrination program as follows:

a. Lectures and films on spatial disorientation are presented (2 to 3 hours on films, 2 hours on lectures).

b. All students participate in a demonstration on the Bárány chair and experience the Coriolis illusion.

c. The best and most frequently used film is the Canadian Air Force "The Perception of Orientation," and the next best is the USAF film "Pilot Vertigo."

d. Two hours of flying unusual attitudes are imposed on the students by the UPT instructors in flight demonstrations occurring halfway in the T-37 phase (instrumentation phase) and again at the beginning of the T-38 phase.

e. Most commonly experienced illusions in flights are listed as: (1) leans; (2) Coriolis; and (3) posturogravic (oculogravic or somatogravic).

f. Most of the students seemed pleased with the present orientation training program.

g. Most students felt that motion in relation to attitude could be learned and that some spatial orientation trainer should be available for such a program, particularly before instrument training.

Physiological training officers (PTOs) expressed a need for a more realistic training program in which a simulator with a capacity of fly-out controls is provided and a need for more research with "operational orientation." Many of them also expressed concern over unfamiliarity with content material involved in spatial orientation indoctrination. By content material they meant the physiological contributing factors involved in spatial orientation and disorientation, and the technological knowledge of aerodynamics in motion stimuli. Since they had never experienced the illusions and were unable to describe them, some PTOs felt that they were not trained sufficiently to instruct or indoctrinate pilots or student pilots on the problems of spatial disorientation.

Two flight-trained PTOs were concerned about the inadequacy of spatial orientation training in that career aircrewmen were not able to use any ground-based demonstrators such as a Bárány chair and that the topic of spatial disorientation was not included in their physiological training refresher course taken every three years. One of these PTOs even claimed that his own flight instructor was "not that familiar with what he was trying to demonstrate" when he undertook in-flight demonstrations of unusual attitude maneuvers.

It is apparent from these interviews that the present requirements for spatial orientation indoctrination are not adequate nor standardized in flight training.

V. RESEARCH STUDIES ON THE SPATIAL ORIENTATION TRAINER (SOT)

Studies have been conducted to determine the SOT's effectiveness in producing sensations and perceptions similar to those imposed by real aircraft during various disorienting flight maneuvers and to estimate its usefulness and acceptability in a flight training program. From the interviews of more than 600 USAF pilots concerning their experiences with flight spatial disorientation, the common flight motion-induced illusions in relating attitude to motion in an aircraft were: leans, posturogravic (oculogravic or "jet upset"), and Coriolis illusions. The SOT is capable of simulating all of these commonly experienced illusions, and the procedure of demonstrating these illusions in the SOT is described elsewhere (19-22).

Twenty-two student pilots in the instrument phase of flight training who were exposed to various disorienting maneuvers in the SOT reported that the SOT (a) increased their confidence in instrument flying as they were able to cope with the takeoff illusion (a common posturogravic illusion) in a real takeoff, (b) increased their faith in the instruments as reporters of real-world conditions and in their ability to respond to them, (c) helped them to maintain the desired attitude by instruments in spite of the illusions experienced (an important condition for a man-in-the-loop trainer), (d) closed the credibility gap by convincing the pilots that their sensations can be at dangerous variance with the true facts, (e) familiarized them with certain disorienting maneuvers that they had not yet experienced in their in-flight unusual-attitude-maneuvers training phase, and (f) afforded them the unrestricted opportunity to work out solutions and make corrections by instruments to motion-induced orientation difficulties that they were not able to

do in actual flight for fear of violating the aircraft structural limitations or the allotment of air space, fuel, time, or altitude (20).

Student pilots, who are not expected to be fully conversant with the bodily sensations induced by flight which are familiar to pilots with long histories of instrumented flight, learned to perform successfully. They responded to different SOT induced illusions by relating these illusions to those they experienced in flight. This indicates a valuable point for such a trainer--the students were capable of vivid perceptual responses to motion because there is a great deal of commonality between the SOT and the aircraft as far as orientation mechanisms are concerned (20, 22).

Subjects much less qualified than student pilots have been able to operate the SOT on relevant illusions in a simulated situation confounded by other illusions. Cramer and Wolfe's study (23) showed that USAF basic trainees experienced the appropriate perceptual responses to changes in pitch motion and accurately solved an orientation problem: they adjusted pitch setting without being diverted by the irrelevant illusion of roll. A later study (24) confirmed the ability of such nonflying personnel to correct for pitch changes even under alcoholic aftereffects. Both studies stressed an important point in simulator training--that man is capable of learning attitude changes in relation to motion.

A different study by Wolfe and Cramer (25) illustrated an important concept in that even sophisticated pilots were shown to be highly susceptible to illusions of pitch during accelerated takeoff or deceleration (posturogravic or somatogravic or jet upset). This study showed that the experiences high-performance aircraft pilots must reintegrate otolithic sensations with visual information from their instrumentation to "cancel out" such disorienting illusions of pitch.

In a study assessing the SOT (21), comparisons were made of other selected orientation training devices in relation to U.S. Air Force flight training in the T-38 airplane. Assessment categories were: motion parameters (yaw, roll, pitch, and Gx axis acceleration); illusions that each device can generate (graveyard spin, Coriolis, leans, posturogravic, jet upset, inversion, elevator effect, autokinesis, oculogyration, false horizontal cues of clouds and lights); type of instrumentation available to the pilot-subject in relation to attitude and control; reaction capability provided in each device for the subject to recover or correct his attitude; reproducible capability to measure and evaluate the subject's performance in coping with disorienting maneuvers; invulnerability to weather; safety; and initial and operating costs. The comparisons indicated that the SOT far excelled other ground-based devices for both hardware- and pilot-oriented assessment criteria. This study emphasized the similarity and reproducibility of SOT-induced illusions to those experienced in flight and the positive feedback to the subject in his selfrecovery by using instruments. The training concept of the SOT was stated in terms of transfer of learning--that is, that senses can be trained to relate attitude to motion and, through practice on a SOT, sensory conflict can be minimized by instrument feedback.

Finally, an evaluation study was made with six pilot instructors in undergraduate pilot training. All six pilot instructors recommended this type of trainer to be used before a student enters the flight instrument phase of training. All six instructors liked the disorientation "symptoms" induced by the SOT.

Table II shows approval of the SOT by various Air Force groups who had been exposed to its use as an aid to spatial orientation training.

VI. OPERATIONAL EXAMPLE OF THE SOT

When a pilot takes off in a high performance aircraft, he experiences an exaggerated sensation of being pitched upward--a common distortion of verticality as perceived by the pilot. If he reacts to this sensation without properly scanning his instruments, he is apt to put his craft into a full power crash. This posturogravic or somatogravic illusion can readily be achieved in the SOT by having the pilot establish his perceptual straight and level condition without using his instruments on the SOT panel. To illustrate this effect, the SOT cockpit is turned inward and is rotated about the main yaw axis to produce a centripetal acceleration of approximately 1 g. When told by the instructor to make the proper adjustments for straight and level, the pilot automatically pitches the cockpit to a nose-down position--as much as 20 degrees. Then, the instructor informs him of his error by allowing him to read the attitude indicator. The pilot continues to practice this maneuver and eventually learns where straight and level flight is. Thus, in the training period, the pilot learns to associate different sets of perceptual criteria with different conditions, in this case, with resultant linear acceleration vectors of different magnitudes. Though this may be oversimplifying a difficult and complex learning task, it is one way of demonstrating man's capacity to cope with such motion stimuli. It is just a beginning.

Operational research regarding other flight functions in relation to the SOT is essential. The real justification for using a SOT in USAF flight training is to minimize spatial disorientation accidents and incidents.

VII. A PROPOSED SPATIAL ORIENTATION TRAINING CONCEPT

At present there is no effective way to practice spatial disorientation maneuvers in the undergraduate pilot training, either in flight or in a ground

environment. Indoctrination on the physiological factors and their relationship to physical stimuli that induce spatial disorientation is undoubtedly essential for training student pilots, pilots, and navigators. A more effective indoctrination program in spatial orientation is necessary to familiarize and train the pilot (1) in identifying and recognizing certain unusual attitude maneuvers, and (2) in convincing the pilot to check his instruments when his sensory information may be in conflict or in error with the actual flight maneuver.

Devices developed for research in spatial disorientation at the USAF School of Aerospace Medicine are illustrated in Figure 1. The SOT is the present climax of this effort. To my knowledge the SOT is the only available closed-loop device developed solely to simulate motion-induced spatial-disorienting maneuvers and provide motion cues needed by the pilot if he is to successfully develop spatial orientation skills.

The block diagram in Figure 2 is necessarily incomplete. It is presented solely to demonstrate: (1) the pervasiveness of the effects of stimuli on the various senses involved in orientation; (2) the complexity of the mechanisms serving instrumented flight; and (3) the sensory systems' input to the overall central nervous system's perception of motion to attitude.

In this diagram the central nervous system (CNS) can take advantage of instrument information in order to correct or adjust the sensory information input. This continuous flow of effects and information between sensory systems and environmental stimuli ultimately accounts for successful spatial orientation, even in an environment of intense stressful modifications (false visual cues, accelerations, gravity, etc.). Both instrument and sensory feedback, and the capacity to respond to it are essential.

The diagram in Figure 3 attempts to present the systems' interfaces in spatial orientation in both the aircraft and a SOT-type simulator. It is

hypothesized that positive transfer of training will occur from a SOT to an airplane since the learning processes of relating motion cues to attitude is similar in both.

In analyzing the pilot's spatial orientation processes, we find ourselves confronted with a series of major questions. What influences, internal or external, act upon man's perceptual system and how are such influences communicated? As depicted here, all this is still at a very crude level of formulation. It will be the task of future researchers in flight training and in spatial orientation training to refine these relationships.

Figure 4 illustrates the methods and aids presently used in spatial orientation indoctrination, as described earlier in this report. Using a SOT-type simulator as a training device for sensory adaptation and learning to relate attitude to motion cues, as well as demonstrating and familiarizing the student pilot with the common motion- and visually-induced illusions in flight, is the innovative approach recommended here.

Studies of spatial disorientation are overwhelmingly of the type in which the subject is a passive transducer of a relatively simple vestibular, proprioceptive, or visual stimuli (26-32). There are very few experiments in which the subject is part of a closed-loop; i.e., his responses exert some influence upon the stimuli to which he is responding, which is the situation of a pilot trying to maintain proper spatial orientation in his aircraft (33-36). Even these few studies are somewhat limited in that they are essentially one-dimensional; i.e., activity is confined to one axis alone. They say nothing, for instance, of judgment of pitch or roll attitude confounded by illusions of pitch or roll induced by linear accelerations as in takeoff, or by Coriolis accelerations induced by noise abatement turns immediately after takeoff, or by combinations of both of these illusions.

In addition, these studies provide no information of the susceptibility of the orientation mechanisms to training. The SOT provides students not only an adequate visual input by way of his instruments, but gives them proprioceptive and vestibular stimulations to simulate flight (19). Since the appropriate stimuli for the vestibular and proprioceptive systems are position and its successive derivatives with respect to time, viz., velocity and acceleration, it is manifestly impossible to reproduce exactly the stimuli generated by an airplane. As a result, a considerable amount of abstraction from the real flight situation must be performed. One must know how a particular orientation problem "looks" to the vestibule. For example, how can the same flight orientation problem be presented to the vestibule in ground-based operations? How closely does the SOT approximate a flight illusion in order to inform the student pilot what to expect in actual flight and how it differs from what is presented to him in the SOT?

In an emergency situation, man generally reacts in accordance with his earliest similar experience. This fundamental rule of learning explains why it is safer to teach the use of aircraft instruments right from the start.

Researchers agree that learning transference can occur (37-39). Since learning is considered a developmental process, prior learning plays an important role in transfer: new learning tasks are placed within the context of habits already established from previous experience. Thus, a student pilot learning a new situation relies on the bank of appropriate habits already mastered. Practice in the SOT will help the pilot control the aircraft in accordance with what he learns from his flight instruments, rather than handicap him by a generalized nonselective switching-off of his sensory system's input. Presumably, through practice on motion simulating devices, proper

processing within the central nervous system of combined inputs from various sensory modalities should reduce the number of conditions which different sensory receptors may be signaling, and assist the pilot in correctly identifying and responding to those conditions imposed upon him by flight (23-25, 40, 41, 44).

A basic concept in learning and in perception is that man uses all sensory modalities to orient himself in space; he does not ignore any of them. The pilot's central nervous system seeks as much information as possible from the senses in order to accurately perceive his own spatial orientation and that of his craft (23-25). External instrument cues help the pilot correct or interpret some sensory information that may conflict with his or his craft's true orientation (41, 44).

Profound changes occur in the pilot as he learns to fly by instruments. He becomes more proficient in decoding information from his instruments as he practices either in aircraft or in static flight simulators. The fate of the input from his vestibule is still a matter of debate. Some argue that the pilot learns to inhibit his vestibular mechanisms so that the input itself is attenuated. A more realistic theory, however, is that reduction of vestibular responses occurs with repeated stimulation as a result of central changes since even the peripheral inhibitory system is subject to central control (23-25). There is also a possibility that the pilot learns to more correctly interpret the information that he receives from his vestibule, so that he operates the controls with good integration of information from his instruments and from his vestibular-kinesthetic system.

This philosophy of distrusting motion cues and suppressing vestibular and proprioceptive sensations is basically against the generally well-accepted rule that learning, retention, and motivation are all improved by involving

as many of the trainee's sense modalities as possible (40, 41). No pilot ignores his sensations, but he uses them to draw his attention to his instruments to determine where his aircraft is (40-44).

In our studies of student pilots, pilot instructors, and advanced jet pilots, we found that they can readily learn to recognize and identify various aspects of their sensations of position and its derivations (20, 23-25). Many of the pilots relate certain SOT motions to flight maneuvers in their commentaries while operating the SOT controls. Pilots also state that visual displays of lights such as ground lights, and star lights, and cockpit and instrument-display lights, are needed to induce some of the visual illusions that they encounter in night flying. They also mention that the cockpit in the SOT should be as similar as possible in configuration to an aircraft, preferably the T-38.

Pilots claim that one of their problems is to successfully perform within a disorienting maneuver. They need more realistic training of unusual attitude maneuvers on a motion system ground-based device that will (1) simulate as accurately as possible the positional disorienting sensations they experience in flight, and (2) have instruments to indicate what they are recovering from, or by which they can correct an attitude. In flight, pilots do not have the time nor the capability, regardless of the safety problem, to practice such recovery techniques.

Thorough research is recommended of a spatial orientation training program, providing two to three hours of instruction in spatial orientation, for use before the instrument flight training phase of the UPT program. This program should include a dialogue between instructor and the student regarding the physiological systems that mediate illusions. Then these

physiologically induced illusions should be presented within the context of flying in the SOT, so that the student can recognize them and interpret them for what they are. Using his instruments the student will operate the SOT controls and then will be reinforced or criticized upon his performance in recovering from or in coping with disorienting maneuvers within the SOT.

In summary, the main training concepts of using the SOT for spatial orientation in flight training are: (1) the basic training function is to convince the pilot that his sensations can vary dangerously from the true facts -- attacking the credibility gap; (2) vestibular habituation or suppression results from repeated vestibular stimulation in that vestibular and proprioceptive sensory inputs from motion are suppressed, but this is a rather tenous suppression; (3) vestibular and proprioceptive senses can be formative if used correctly through training and reinforcement on a SOT-type simulator.

VIII. CONCLUSIONS

Studies using the SOT indicate that pilots use their vestibular and proprioceptive senses as formative instruments rather than ignore or suppress these senses (23-25, 35, 41), which is the natural way for man to perceive and correct for any sensory conflict (41). The pilot can correct sensory conflict by identifying the problem through familiarization or experience, and by cross-checking external cues such as attitude, motion, and directional instruments to help him orient himself (41, 44). Resistance to disturbing and disorienting maneuvers is also increased by familiarization and practice with such stimuli (28, 34, 35).

The SOT has been valuable in showing the need for a closed-loop motion trainer to help a pilot learn to recover from and cope with spatial disorienting

maneuvers. The staff at USAFSAM who have investigated the problem of spatial orientation/disorientation training for the past decade have found that the trainee must receive positive feedback of the true position of the trainer and what illusion the trainer is simulating.

The validity of any training program for spatial orientation will be established by comparing functions obtained in the trainer with those from the same population in the aircraft (41). Based on design studies and cross-validation of learning transference, a safe and economical means of providing pretraining spatial disorientation maneuvers and corrective tasks is feasible. The transfer-or-training potential of this SOT, as illustrated in Figures 3 and 4, should significantly reduce the actual flight time required for the pilot to reach desired levels of in-flight proficiency or to cope with a spatial disorienting maneuver.

Along with using a SOT within a spatial orientation training syllabus, we recommend supplementary material on the senses involved, keyed to demonstration and training within a SOT-type simulator. Together, they can help the student pilot overcome sensory conflicts that result in disorientation and/or practice proper aircraft control in spite of disorienting sensory cues.

REFERENCES

1. Lewis, S. T., H. Whitmore, L. Harris, and D. McDougall. A spatial disorientation demonstrator. SAM-TDR-65-7, March 1965.

2. Doppelt, F. F. A program for the spatial disorientation demonstrator. SAM-TR-65-66, October 1965.

3. Luehrs, R. Significance of vertigo as a cause of accidents. Presented at the Symposium on Vertigo as a Problem in Aerospace Medicine, U.S. Naval Air Development Center, Johnsville, Pa., 1967.

4. Smiley, J. Aeromedical incidents amongst Canadian forces pilots: A survey. Presented at the 38th Annual Meeting of the Aerospace Medical Association, Washington, D. C., 1967.

5. AGARD. Conference reprint No. 95 on "The Disorientation Incident." North Atlantic Treaty Organization, September 1971.

6. Barnum, F. Spatial Disorientation aircraft accidents, 1 January 1969-31 December 1971, A comparison. Presented at the 44th Annual Meeting of the Aerospace Medical Association, Las Vegas, Nevada, 1973.

7. Barnum, F., and R. H. Bonner. The Epidemiology of USAF spatial disorientation aircraft accidents, 1 January 1958-31 December 1968. Aerosp. Med. 42:896-989 (1971).

8. Moser, R., Jr. Spatial disorientation as a factor in accidents in an operational command. Aerosp. Med. 40:174-176 (1969).

9. Nuttall, J. B. The problem of spatial disorientation. JAMA 166:431-438 (1958).

10. Neely, S., A. Zeller, and G. H. Normand. Indicated problems and reasons for conference-spatial disorientation. In Proceedings of the First Conference on Vestibular Physiology and Spatial Disorientation. USAFSAM, June 1958.

11. Hixson, W. C., J. I. Niven, and E. Spezia. Orientation-error accidents in Army aviation aircrafts. AGARD, conference preprint No. 95, "The Disorientating Incident," AGARD-CPT 95-71, 1971.

12. Niven, J. I., W. C. Hixson, and E. Spezia. Orientation-error accidents in regular Army aircraft during fiscal year 1968: relative incidents and cost. U.S. Army Aeromedical Laboratory and Naval Aerospace Research Laboratory Joint Report. NAMRL-1143, September 1971.

13. Hixson, W. C., J. I. Niven, and E. Spezia. Orientation-error accidents in regular Army aircraft during fiscal year 1969: Relative incidence and cost. U.S. Army Aeromedical Research Laboratory and Naval Aerospace Medical Research Laboratory Joint Report. NAMRL-1161, April 1972.

14. Tyler, P. E., and P. A. Furr. Disorientation, "Fact or Fancy." AGARD conference preprint No. 95 "The Disorientation Incident," AGARD CPT-95-71, 1971.

15. Lofting, R. G. A review of United Kingdom (RAF and Army) Statistics on Spatial Disorientation in Flight, 1960-1970. AGARD Conference preprint No. 95, "The Disorientation Incident" AGARD-CPT-95-71, 1971.

16. Ninow, E. H., W. F. Cunningham, and F. A. Radcliffe. Psychophysiological and environmental factors affecting disorientation in Naval aircraft accidents. AGARD conference preprint No. 95, "The Disorientation Incident," AGARD-CPT-95-71, 1971.

17. Clark, B. Disorientation incidents reported by military pilots across 14 years of flight. AGARD conference preprint No. 95, "The Disorientation Incident," AGARD-CPT-95-71, 1971.

18. Rayman, R. B. Sudden incapacitation in flight, 1 January 1966-30 November 1971. Aerosp. Med. 44:953-955 (1973).

19. Gillingham, K. K., R. L. Cramer, P. J. Dowd, and F. G. Collins. Design criteria for the spatial orientation trainer. SAM-TR-67-90, October 1967.

20. Dowd, P. J., R. L. Cramer, J. W. Wolfe, and S. H. McKean, III. Responses of USAF Undergraduate Pilot Trainees to indoctrination in the spatial orientation trainer. Aerosp. Med. 41:544-549 (1970).

21. Dowd, P. J. A critical assessment of ground-based devices for spatial orientation training. USAFSAM-TR-73-23, August, 1973.

22. Dowd, P. J. The USAFSAM spatial orientation trainer: Background and apparatus. USAFSAM-TR (in press).

23. Cramer, R. L., and J. W. Wolfe. Effects of pitch and Coriolis illusions upon adjustment of pitch angle. Aerosp. Med. 41:644-646, (1970).

24. Dowd, P. J., J. W. Wolfe, and R. L. Cramer. Aftereffects of alcohol on the perception and control of pitch-attitude during centripetal acceleration. Aerosp. Med. 44:928-930 (1973).

25. Wolfe, J. W., and R. L. Cramer. Illusions of pitch induced by centripetal acceleration. Aerosp. Med. 41:1136-1139 (1970).

26. Benson, A. J. Spatial disorientation in flight. A textbook of Aviation Physiology. Ed: J. A. Gillies, New York: Pergamon Press, pp. 1086-1129, 1965.

27. Brandt, V. Reorientation and vestibular function. Acta Oto-laryngol. 50:163-170 (1959).

28. Dowd, P. J., and R. L. Cramer. Habituation transference in Coriolis acceleration. Aerosp. Med. 38:1103-1107, (1967).

29. Clark, B. Visual space perception as influenced by unusual vestibular stimulation. Human Factors, 5:265-274 (1963).

30. Graybiel, A. Orientation in aerospace flight. Spec. Rept. 66-6, Naval Aerospace Medical Institute, October 1966.

31. Guedry, F. E., Jr., and E. K. Montague. Quantitative evaluation of the vestibular Coriolos reaction. Aerosp. Med. 32:487-500 (1961).

32. Stewart, J. D., and B. Clark. Coriolis effects during pitch and roll maneuvers in a piloted flight simulator. Aerosp. Med. 36:105-112 (1965).

33. Collins, W. B. Coriolis vestibular stimulation and the influence of different visual surrounds. Aerosp. Med. 39:125-130 (1968).

34. Gillingham, K. K. Some notes on the threshold of the vestibular Coriolis effect and its significance to aircrew. SAM-TR-65-55, August 1965.

35. Gillingham, K. K. Training the vestibule for aerospace operations: Using Coriolis effect to assess rotation. Aerosp. Med. 37:47-51 (1966).

36. Graybiel, A. and B. Clark. Validity of the oculogravic illusion as a specific indicator of otolith function. Aerosp. Med. 36:1173-1181 (1965).

37. Ferguson, G. A. Transfer and human ability. Canad. J. Psychol., 10:121-131 (1956).

38. Fine, P. N. and C. L. Jennings. Coping and developmental theory: Applicability to selective study of normal men. USAFSAM Rev. 1-65, April 1965.

39. Wood, M. E., W. V. Hagin, O. O'Conner and D. W. Myers. Design of a simplified formation trainer. AFHRL-TR-72-8, March 1972.

40. Borlace, F. H. Flight simulator motion, its enhancement and potential for flight crew training. Presented at the Society of Automotive Engineers Third International Simulation and Training Conference, 1967.

41. Gibino, D. J. Effects of presence or absence of cockpit motion in instrument flight trainers and flight simulators. Aeronautical Systems Div., AFSC, Wright-Patterson AFB, Ohio, ASD-TR-68-24, June 1968.

42. Matheny, W. G., D. J. Dougherty, and J. M. Willis. Relative motion or elements in instrument displays. Aerosp. Med. 34:1041-1046 (1963).

43. Huddleston, H. F. Cockpit motion requirements for flight simulation. The RAF Institute of Aviation, Farnborough, England IAM Report No. 363, January 1966.

44. Puig, J. A. Motion in flight training: A human factors view. Naval Training Device Center, Orlando, Florida, NAVTRADEVCEN IH-177, October 1970.

45. Munson, H. G., Maj., USAF, MC, AFISC/SEL, Norton AFB CA-Personal Communication.

TABLE I. SUMMARY OF AIRCRAFT ACCIDENTS

$ Cost/yr	ACCIDENTS Total		SD ACCIDENTS Total SD			FATAL ACCIDENTS Total		SD FATAL ACCIDENTS Total SD			Organization	Year	Source
	N	N/yr	N	N/yr	%	N	N/yr	N	N/yr	%			
	640	213	39	13	6	239	80	24	8	10	USAF	1969-71	(6)
			21	21				8	8		USAF	1971	(45)
	4679	425	281	26	6	1462	133	211	19	15	USAF	1958-69	(7)
					9					26	USAF	1964-67	(8)
18.3 M	3764	1501	76	31	2	1892	756	90	36	5	USAF	1955-58	(10)
					4					14	USAF	1955-56	(9)
11.9 M	701	701	71	71	6.2	346	346	22	22	13.02	USA	1969	(13)
12.4 M	702	702	75	75	7.2	232	232	75	75	18.84	USA	1968	(12)
10.1 M	802	802	57	57	7.1	124	124	19	19	15.32	USA	1967	(11)
			102	51							USN	1969-70	(16)
	712	712	48	48	6.75						USN	1969	(14)
			29	3							RAF	1960-70	(15)

SD - spatial disorientation

% - percent of total accidents

M - millions

TABLE I. SUMMARY OF AIRCRAFT ACCIDENTS

$ Cost/yr	ACCIDENTS Total N	ACCIDENTS Total N/yr	SD ACCIDENTS Total SD N	SD ACCIDENTS Total SD N/yr	SD ACCIDENTS Total SD %	FATAL ACCIDENTS Total N	FATAL ACCIDENTS Total N/yr	SD FATAL ACCIDENTS Total SD N	SD FATAL ACCIDENTS Total SD N/yr	SD FATAL ACCIDENTS Total SD %	Organization	Year	Source
	640	213	39	13	6	239	80	24	8	10	USAF	1969-71	(6)
			21	21				8	8		USAF	1971	(45)
	4679	425	281	26	6	1462	133	211	19	15	USAF	1958-69	(7)
					9					26	USAF	1964-67	(8)
18.3 M	3764	1501	76	31	2	1892	756	90	36	5	USAF	1955-58	(10)
					4					14	USAF	1955-56	(9)
11.9 M	701	701	71	71	6.2	346	346	22	22	13.02	USA	1969	(13)
12.4 M	702	702	75	75	7.2	232	232	75	75	18.84	USA	1968	(12)
10.1 M	802	802	57	57	7.1	124	124	19	19	15.32	USA	1967	(11)
			102	51							USN	1969-70	(16)
	712	712	48	48	6.75						USN	1969	(14)
			29	3							RAF	1960-70	(15)

SD - spatial disorientation

% - percent of total accidents

M - millions

TABLE II

APPLICATION OF THE SOT AS AN AID IN SPATIAL ORIENTATION TRAINING

	Yes	No
PTO's	3	--
Students	21	1
Instructors (UPT)	6	--
Jet Pilots	17	--

LEGENDS

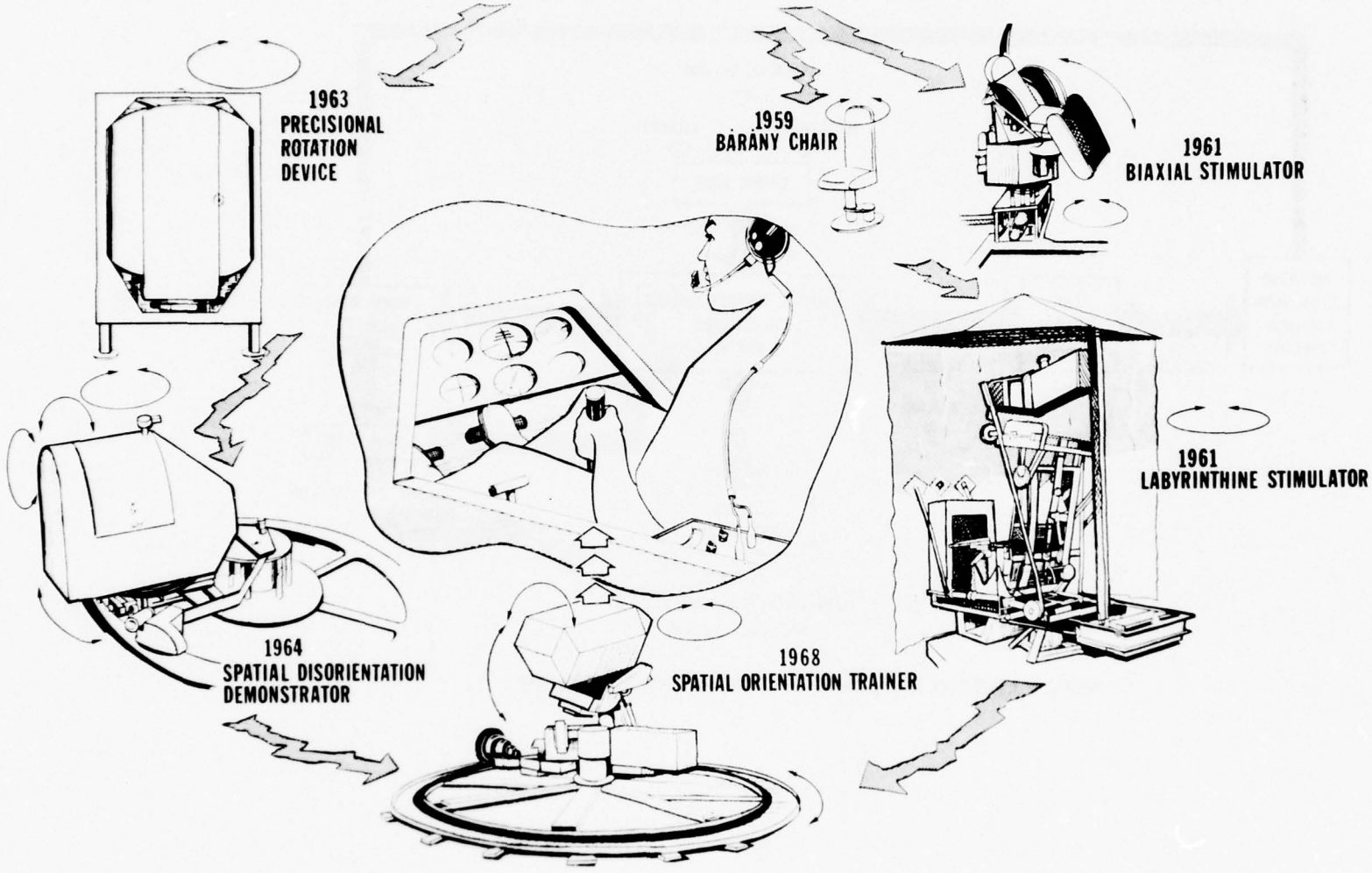
Figure 1. USAFSAM Ground-Based Vestibular Stimulators.

Figure 2. Servomechanisms in Spatial Orientation.

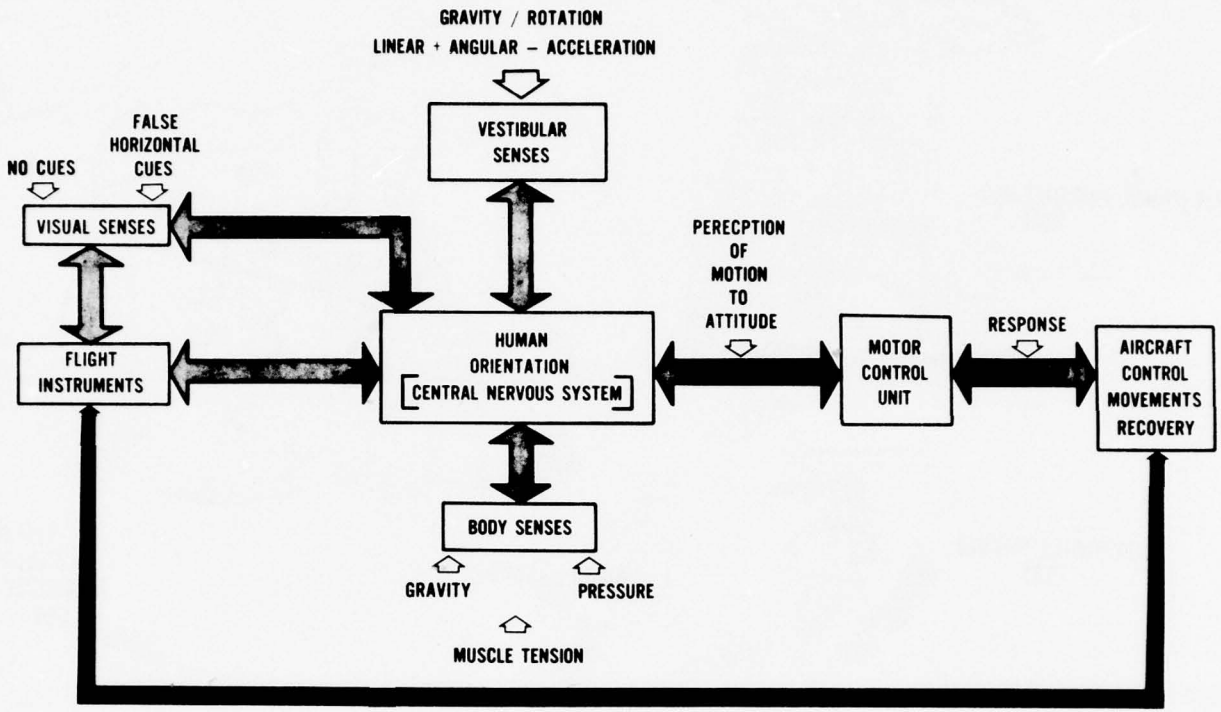
Figure 3. Systems Interfaces in Spatial Orientation.

Figure 4. Spatial Disorientation Indoctrination.

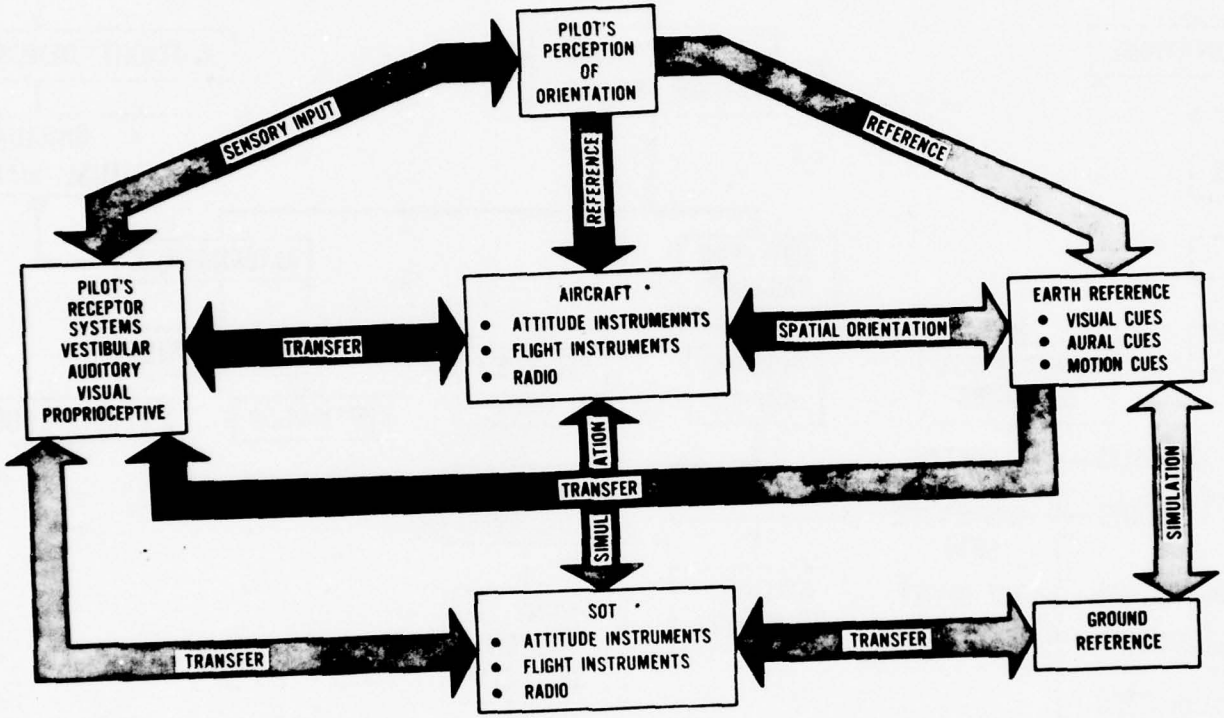
USAFSAM
GROUND-BASED VESTIBULAR STIMULATORS



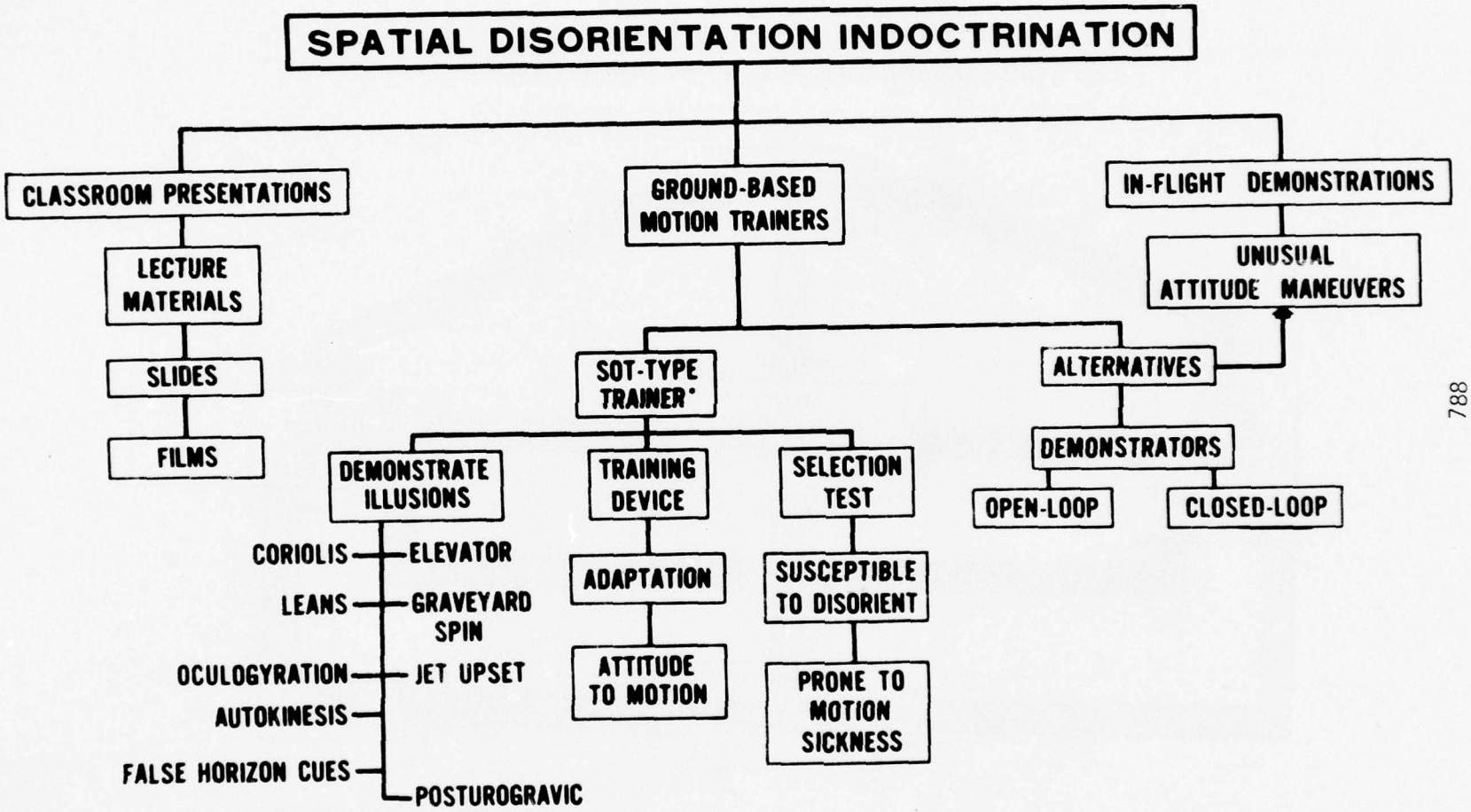
SERVOMECHANISMS IN SPATIAL ORIENTATION



SYSTEMS INTERFACES IN SPATIAL ORIENTATION



* TRUE INDICATION OF POSITION OF COCKPIT IN RELATION TO EARTH



788

* CLOSED-LOOP MOTION SIMULATOR

MINUTES

1973 MTA Steering Committee Meeting

The 1973 Military Testing Association (MTA) Steering Committee Meeting convened at San Antonio, Texas at 1330, 30 October 1973. The following members were present:

U. S. Air Force
(Coordinating Agency)

Col Theodore B. Aldrich,
President
Dr. L. D. Brokaw, Secretary
Col F. L. McLanathan, Retired
Mr. L. N. Jaimes

U. S. Navy

Cdr M. W. Faessel
Mr. C. J. Macaluso

U. S. Army

Col W. F. Veaudry
Dr. E. R. Dusek
Mr. W. W. Wance
Mr. Anthony Strigari

U. S. Marine Corps

Maj H. E. Wells
Mr Edward A. Dover

U. S. Coast Guard

Cdr K. R. Depperman
Mr. John A. Burt
Cdr K. M. Roughgarden

Australia

Lt Cdr J. E. C. Williams, RAN
Sqn Ldr W. J. Watson, RAAF

Canada

Cdr W. H. Northey, CF Hq

The meeting was called to order by the MTA President who presented the following Agenda items to be discussed: By-Laws, the 1974 MTA meeting, organizational name change, and miscellaneous items requiring committee action.

1. Mr. Jaimes, AFHRL, opened the disucssion regarding the funding of MTA meetings outlining the AFR requirements as ennumerated in AFR 80-43. Discussion was invited regarding the fee versus a registration fee for funding purposes. Under AFR 80-43, an Air Force organization hosting a meeting or symposium is expected to pay registration fees for all DoD participants. A coordinator on the other hand pays no such fees. The present meeting is being funded by a very small amount of AF funds plus a registration fee imposed on each conferee. The registration fee this year will include the cost of printing the MTA Proceedings. A copy of the proceedings will be forwarded to each attendee.

2. Mr. John Burt, USCG, made a motion to change the term "host" in the current By-Laws to "coordinating." Cdr Roughgarden, USCG, seconded the motion which carried with no opposition.

3. Dr. Brokaw, AFHRL, moved that the name of the MTA should be changed. This motion was seconded by Mr. Anthony Strigari, USA, and the final vote was 5 to 4 in favor of the proposed action. A further motion was made by Dr. Brokaw concerning the method to be used in effecting a name change. Dr. Brokaw will circulate correspondence to Steering Committee members seeking suggestions for name changes. When they are returned, if there is a consensus, he will prepare a ballot which will be distributed to the membership, defined as registrants to this year's conference, for their completion. Ballots will then be returned to the Coast Guard in Oklahoma City, and results will be reported at next year's MTA meeting. The motion was seconded by Mr. Strigari, USA, and passed unanimously.

4. Upon a query by the U. S. Coast Guard concerning the cost of the MTA, Mr. Jaimes, AFHRL, stated that he will prepare a financial statement, and a transfer of remaining funds will be made to the Coast Guard. In addition, a transfer of the MTST tape files used in mail-out will be made available to the Coast Guard.

5. Dr. Brokaw, in order to simplify procedures for the program chairman of the 74 conference, offers the following suggestions to the Coast Guard:

a. In your call for the papers, provide a form to be completed somewhat like that of the American Psychological Association or of the Air Force Academy. On this form, the would-be participant must indicate the overall time for his paper plus discussion, he must acknowledge awareness of a requirement for photo reproducible copy to be used in the proceedings, and careful specification of that copy should be made. For example, single spaced with 1 1/4 inch margins on either side, 1 inch on top and 3/4 inch on the bottom, 8 by 10 1/2 paper. It must be taped with either a carbon tape or with a ribbon designed for photo offset reproduction, etc.

b. The Air Force Academy uses a procedure which could be very useful in terms of expediting the publication of the proceedings and also the clearance of papers. The Air Force Academy requires a reproducible copy be delivered to them not less than 30 days prior to the opening day of their conference with indication to the participant that his paper will be withheld if clearance is not forthcoming.

6. We had logistic and financial problems which prevented our doing so, but there was considerable evidence that we might have had an even larger turnout if we had been able to furnish a tentative program to our mailing list.

7. It is probably inevitable that you will have a few last minute cancellations and it has been suggested that various means might be taken to counter such changes. For example, persons in the local area who have a paper ready for presentation and who are amenable to "filling in" could be used. We had about three of those. However, I see no way to avoid the somewhat unfair treatment of those persons who would have participated except their papers had been rejected.

8. Commander Northey, CDF, suggested that the concluding remarks by the MTA President solicit comments from the MTA participants as to whether the single-session, one-room, or a multiple-track meeting with two or three sessions operating simultaneously was preferred for the next meeting.

9. Colonel F. L. McLanathan (Retired) moved that a citation be awarded to an outstanding MTA member on an annual basis in the name of Captain Harry Greer, U. S. Navy (Deceased) and presented at the MTA banquet. Commander Northey, CDF, seconded the motion. The following procedure for nomination was outlined. Nominations for the award will be submitted to the Steering Committee for consideration and final selection. The responsibility for the award and accompanying citation will be the responsibility of the coordinating agency for that meeting. This motion unanimously carried.

Dr. Brokaw, MTA Secretary (AFHRL), moved to adjourn. C. J. Macaluso, USN, seconded and the meeting was adjourned.

MILITARY TESTING ASSOCIATION

HARRY H. GREER AWARD

In the memory of Captain Harry H. Greer, USN, founder of the Military Testing Association, the first Harry H. Greer award is hereby presented to Charles J. Macaluso, United States Naval Examining Center, for his outstanding services to the organization. From its inception, the Military Testing Association has been indebted to this man.

Through foresight, devotion, and dedication, he has made a substantial contribution to raising the stature of the association from a small discussion group to a major international organization which now provides a forum for specialists in many areas of personnel evaluation. Charles J. Macaluso was instrumental in planning the first meeting of the association in 1959 and each year since he has been a major factor in the success of its conventions. He has devoted countless hours to furthering the aims of the association across the Department of Defense and in other nations. His zeal and energy in developing the objectives of the association have been an example and inspiration to all its members. It is with gratitude and appreciation that the Military Testing Association recognizes the efforts of Charles J. Macaluso with the Harry H. Greer award.

PRECEDING PAGE BLANK-NOT FILMED

BY-LAWS OF THE MILITARY TESTING ASSOCIATION

Article I - Name

The name of this organization shall be the Military Testing Association.

Article II - Purpose

The purpose of this Association shall be to:

A. Assemble representatives of the various armed services of the United States and such other nations as might request to discuss and exchange ideas concerning assessment of military personnel.

B. Review, study, and discuss the mission, organization, operations, and research activities of the various associated organizations engaged in military personnel assessment.

C. Foster improved personnel assessment through exploration and presentation of new techniques and procedures for behavioral measurement, occupational analysis, manpower analysis, simulation models, training programs, selection methodology, survey and feedback systems.

D. Promote cooperation in the exchange of assessment procedures, techniques and instruments.

E. Stimulate and coordinate research in the various areas of military personnel systems.

F. Promote the assessment of military personnel as a scientific adjunct to modern military personnel management within the military and professional communities.

Article III - Participation

The following categories shall constitute membership within the MTA:

A. Primary Membership.

1. All active duty military and civilian personnel permanently assigned to an agency of the associated armed services having primary responsibility for assessment for personnel systems;

2. All civilian and active duty military personnel permanently assigned to an organization exercising direct command over an agency of the associated armed services holding primary responsibility for assessment of military personnel.

PRECEDING PAGE BLANK-NOT FILMED

B. Associate Membership.

1. Membership in this category will be extended to permanent personnel of various governmental, educational, business, industrial and private organizations engaged in activities that parallel those of the primary membership. Associate members shall be entitled to all privileges of primary members with the exception of membership on the Steering Committee. This restriction may be waived by the majority vote of the Steering Committee.

Article IV - Dues

No annual dues shall be levied against the participants.

Article V - Steering Committee

A. The governing body of the Association shall be the Steering Committee. The Steering Committee shall consist of voting and non-voting members. Voting members are primary members of the Steering Committee. Primary membership shall include:

1. The Commanding Officers of the respective agencies of the armed services exercising responsibility for personnel assessment programs.

2. The ranking civilian professional employees of the respective agencies of the armed service exercising primary responsibility for the conduct of personnel assessment systems. Each agency shall have no more than two (2) professional civilian representatives.

B. Associate membership of the Steering Committee shall be extended by majority vote of the committee to representatives of various governmental, educational, business, industrial and private organizations whose purposes parallel those of the Association.

C. The Chairman of the Steering Committee shall be selected through election by the majority vote of the committee. The term of office shall be one year and shall begin the last day of the annual conference.

D. The Steering Committee shall have general supervision over the affairs of the Association and shall have the responsibility for all activities of the Association. The Steering Committee shall conduct the business of the Association in the interim between annual conferences of the Association by such means of communication as deemed appropriate by the President or Chairman.

E. Meeting of the Steering Committee shall be held during the annual conferences of the Association and at such times as requested by the President of the Association or the Chairman of the Steering Committee. A majority of the members of the Steering Committee shall constitute a quorum.

Article VI - Officers

A. The officers of the association shall consist of a President, Chairman of the Steering Committee and a Secretary.

B. The President of the Association shall be the Commanding Officer of the armed services agency coordinating the annual conference of the Association. The term of the President shall begin at the close of the annual conference of the Association and shall expire at the close of the next annual conference.

C. It shall be the duty of the President to organize and coordinate the annual conference of the Association held during his term of office, and to perform the customary duties of a president.

D. The Secretary of the Association shall be filled through appointment by the President of the Association. The term of office of the Secretary shall be the same as that of the President.

E. It shall be the duty of the Secretary of the Association to keep the records of the association, and the Steering Committee, and to conduct official correspondence of the association, and to issue notices for conferences. The Secretary shall also perform such additional duties and take such additional responsibilities as the President may delegate to him.

Article VII - Meetings

A. The Association shall hold a conference annually.

B. The annual conference of the Association shall be coordinated by the agencies of the associated armed services exercising primary responsibility for military personnel assessment in order of the following rotating schedule:

United States Army
United States Marine Corps
United States Navy
United States Air Force
United States Coast Guard

C. The annual conference of the Association shall be held at a time and place determined by the coordinating agency. The membership of the association shall be informed at the annual conference of the place at which the following annual conference will be held. The coordinating agency shall inform the Steering Committee of the time of the annual conference not less than six (6) months prior to the conference.

D. The coordinating agency shall exercise planning and supervision over the program of the annual conference. Final selection of program content shall be the responsibility of the coordinating organization.

E. Any other organization desiring to coordinate the conference may submit a formal request to the Chairman of the Steering Committee, no later than 18 months prior to the date they wish to serve as host.

Article VIII - Committees

A. Standing committees may be named from time to time, as required, by vote of the Steering Committee. The chairman of each standing committee shall be appointed by the Chairman of the Steering Committee. Members of standing committees shall be appointed by the Chairman of the Steering Committee in consultation with the chairman of the committee in question. Chairmen and committee members shall serve in their appointed capacities at the discretion of the Chairman of the Steering Committee. The Chairman of the Steering Committee shall be an ex officio member of all standing committees.

B. The President with the counsel and approval of the Steering Committee may appoint such ad hoc committees as are needed from time to time. An ad hoc committee shall serve until its assigned task is completed or for the length of time specified by the President in consultation with the Steering Committee.

C. All standing committees shall clear their general plans of action and new policies through the Steering Committee, and no committee or committee chairman shall enter into relationships or activities with persons or groups outside of the Association that extend beyond the approved general plan of work without the specific authorization of the Steering Committee.

D. In the interest of continuity, if any officer or member has any duty elected or appointed placed on him, and is unable to perform the designated duty, he should decline and notify at once the officers of the association that he cannot accept or continue said duty.

Article IX - Amendments

A. Ammendments of these By-Laws may be made at any annual conference of the Association.

B. Amendments of the By-Laws may be made by majority vote of the assembled membership of the Association provided that the proposed amendments shall have been approved by a majority vote of the Steering Committee.

C. Proposed amendments not approved by a majority vote of the Steering Committee shall require a two-third's vote of the assembled membership of the Association.

Article X - Voting

All members in attendance shall be voting members.

Article XI - Enactment

These By-Laws shall be in force immediately upon acceptance by a majority of the assembled membership of the Association and/or amended (in force 2 November 1973).

LISTING OF CONFEREES

ABELLERA, James W., 1Lt
AFHRL/PE, Lackland AFB, Texas 78236

ALDRICH, Theodore B., Col
AFHRL/MD, Lackland AFB, Texas 78236

ALEXANDER, Helen A.
AFHRL/MD, Lackland AFB, Texas 78236

ALLEN, Eugene D., Capt
Department of Education Evaluation, CE School, AFIT/DEE, Wright Patterson AFB, Ohio 45433

ALLEN, Marlin R., Capt
AFHRL/SM, Lackland AFB, Texas 78236

ALLEY, William E.
AFHRL/PE, Lackland AFB, Texas 78236

ANDERSON, Charles
Educational Testing Service, 960 Grove Street, Evanston, Illinois 60201

ANSBRO, Thomas M.
Service School Command, Naval Training Center, Orlando, Florida 32813

ARCHER, Joann R.
AFHRL/OR, Lackland AFB, Texas 78236

ARCHER, Wayne B.
AFHRL/OR, Lackland AFB, Texas 78236

BADALAMENTE, Richard V., Maj
3714 36th Street, Lubbock, Texas 79413

BAILEY, James T., Sgt
AFHRL/SM, Lackland AFB, Texas 78236

BAKER, Thomas B.
ATSL-TE, USA Ordnance Center & School, Aberdeen Proving Ground, Maryland 21005

BARTON, Bruce B., A1C
AFHRL/SM, Lackland AFB, Texas 78236

BELT, Dr. John A.
Wichita State University, Wichita, Kansas 67208

BERBERICH, George L., Amn
AFHRL/PE, Lackland AFB, Texas 78236

BERGE, Paul
USAFI, 2318 S. Park Street, Madison, Wisconsin 53713

BERGMAN, Joseph
AFHRL/OR, Lackland AFB, Texas 78236

BERTHOLD, Oscar A., Col
AFHRL/CCV, Brooks AFB, Texas 78235

BEST, William J., A1C
AFHRL/SM, Lackland AFB, Texas 78236

BIEDIGER, Rosalie A.
AFHRL/SM, Lackland AFB, Texas 78236

BIRDSALL, Walter
Naval Examining Center, Bldg 2711, Great Lakes, Illinois 60088

BLAIR, Ben R., Jr.
3700 OCMST, Lackland AFB, Texas 78236

PRECEDING PAGE BLANK-NOT FILMED

BOHATCH, Andrew R.
Naval Examining Center Bldg 2711, Great Lakes Naval Training Center, Great Lakes, Illinois

BOLLER, Lawrence J., Maj
Office of Manpower Utilization, HQ MC AOIH, Marine Corps Base, Quantico, Virginia 22134

BONDARUCK, John
National Security Agency, Department of Defense, Ft Meade, Maryland 20730

BOTTENBERG, Dr. Robert A.
AFHRL/SM, Lackland AFB, Texas 78236

BOWSER, S. E.
Navy Personnel Research & Development Center, Point Loma, San Diego, California 92152

BOYD, Dr. Joseph L.
Educational Testing Service, Princeton, New Jersey 08540

BRAUN, F. B.
Data-Design Laboratories, 803 W. Broad Street, Falls Church, Virginia 22046

BRAZEL, James L.
AFHRL/SM, Lackland AFB, Texas 78236

BROKAW, Dr. Leland D.
AFHRL/PE, Lackland AFB, Texas 78236

BROWN, Bruce, Cdr
CNET Staff, Pensacola, Florida 32508

BROWN, Thomas A.
Info Sciences & Math Department, RAND, Santa Monica, California 90406

BRUMBACK, Dr. Gary B.
AIR, 8555 16th Street, Silver Spring, Maryland 20910

BULLOCK, Kay D.
United Services Automobile Association, 4119 Broadway, San Antonio, Texas 78288

BURT, John A.
U. S. Coast Guard Institute, P. O. Substation 18, Oklahoma City, Oklahoma 73169

BUSBY, Charles W., Cdr
U. S. Coast Guard, USCG Training Center, Petaluma, California 94952

CALHOUN, Gary J., AlC
AFHRL/SM, Lackland AFB, Texas 78236

CALLAHAN, James P.
USA Enl Eval Ctr, Ft Benjamin Harrison, Indianapolis, Indiana 46227

CAMPBELL, Bruce A.
Science Research Associates, 259 E. Erie, Chicago, Illinois 60611

CANIPE, John F., Jr., Capt
3700 OCMS, Lackland AFB, Texas 78236

CANTER, Dr. Ralph R.
Director, Manpower Rsch, OSD (Manpower), Rm 3D 986, The Pentagon, Washington, D. C. 20301

CARPENTER, Gerald J., Lt Col
Head, Dept of Military Leadership & Management, Royal Military College of Canada, Kingston, Ontario, Canada

CARPENTER, James B., Lt Col
AFHRL/OR, Lackland AFB, Texas 78236

CHAMBERS, Dr. Randall M.
School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332

CHENEY, Knight D., Lt
Reserve Training Division, US Coast Guard Headquarters, Washington, D. C. 20590

CHRISTAL, Dr. Raymond E.
AFHRL/OR, Lackland AFB, Texas 78236

CHRISTENSEN, James E., Sgt
AFHRL/SM, Lackland AFB, Texas 78236

CLARK, Henry W.
AFHRL/SM, Lackland AFB, Texas 78236

CLARK, Dr. Herbert J.
AFHRL/XP, Brooks AFB, Texas 78235

COHEN, Dr. Edwin
Simulation Products Division, Singer Company, Binghamton, New York 13902

COLE, Elsie M.
AFHRL/DOJZ, Brooks AFB, Texas 78235

COPA, Dr. George
University of Minnesota, 145 Peik Hall, Minneapolis, Minnesota 55455

COPE, Carl G.
Evaluation Division, DRTM, Ft Rucker, Alabama 36360

CORMACK, Bruce, Cmdr
National Defence Headquarters, Ottawa, Ontario Canada K1AOK2

COURTRIGHT, John F., Lt Col
Air University (DPM), Maxwell AFB, Alabama 36112

COWAN, Douglas K.
AFHRL/PE, Lackland AFB, Texas 78236

CRITTENDON, Richard E., MSgt
AFHRL/DAU, Lackland AFB, Texas 78236

CROCHET, Charles R., AlC
AFHRL/SM, Lackland AFB, Texas 78236

CROLL, Paul R.
Sys Dev & Eval Office TAGCEN, Forrestal Bldg, Rm GB-137, Washington, D. C. 20314

CROW, Graham P., Maj
AFHRL/XP, Brooks AFB, Texas 78235

CROW, Ralph, Jr., Maj
Training Research Applications Division, AF Mil Tng Ctr, Lackland AFB, Texas 78236

CUMMINS, Gary J., Capt
OMU (AOIH) HQMC, Marine Corps Base, Quantico, Virginia 22134

CURTO, Salvatore, Capt
HQ USAF/DPXYR, Rm 5C428, Pentagon, Washington, D. C. 20330
DAHLE, Lee J., 1Lt
AFHRL/DOGR, Brooks AFB, Texas 78235
DAMEWOOD, Logan A., Lt Col
AFHRL/DOG, Brooks AFB, Texas 78235
DASH, Roy E., Cdr
USCG Training Center, Cape May, New Jersey 08204
DEENY, Charles B.
Southern Operations Marketing Mgr, National Computer Systems, 5815 Vestavia Drive, Houston, Texas 77069
DEGAUGH, Dr. Roy
AFHRL/MD, Lackland AFB, Texas 78236
DEPPERMAN, K. R., Cdr
U. S. Coast Guard Institute, P. O. Substation 18, Oklahoma City, Oklahoma 73169
DEVANGE, Thomas M., Sgt
AFHRL/PE, Lackland AFB, Texas 78236
DOVER, Edward A.
HQS Marine Corps, Arlington, Virginia
DUNHAM, Alan D., Capt
6970 Support Group/R211, Ft Meade, Maryland 20755
DUSEK, Dr. Edwin R.
US Army Rsch Institute for the Behavioral and Social Sciences, 1300 Wilson Boulevard, Arlington, Virginia 22209
EARLES, James A., Amn
AFHRL/PE, Lackland AFB, Texas 78236
ELSTER, Dr. Richard S.
U. S. Naval Postgraduate School, Monterey, California 93940
FAESSEL, M. W., Cdr
Commanding Officer, Naval Examining Center, Bldg 2711, Great Lakes, Illinois 60088
FIGEL, William J.
Test Development, Science Research Associates, Inc., 259 East Erie Street, Chicago, Illinois 60611
FISCHER, Harold E., Col
AFHRL/CC, Brooks AFB, Texas 78235
FISCHL, Dr. Myron A.
U. S. Army Res Inst, 1300 Wilson Boulevard, Arlington, Virginia 22209
FITES, Jeanne B.
AFHRL/MD, Room 200, 701 Prince Street, Alexandria, Virginia 22314
FLEMING, Jimmy L., 2Lt
Community College of the Air Force, Randolph AFB, Texas 78148
FRESNE, Calvin C.
AFHRL/SM, Lackland AFB, Texas 78236
FRIEMANN, James L.
AFHRL/SM, Lackland AFB, Texas 78236

GAINES, Gerald D., Col
Coordinator of Training, Division of Veterinary Medicine, Walter Reed Institute of Research, Washington, D. C.

GARRITY, John M., Capt
3700 OCMST, Lackland AFB, Texas 78236

GAMBRELL, Melvin B., Jr., Capt
HQ ATC RST, Randolph AFB, Texas 78148

GARZA, Andrew T.
AFHRL/OR, Lackland AFB, Texas 78236

GAY, Lenwood O.
National Security Agency, 9800 Savage Road, Ft Meade, Maryland 20755

GEISEN, John A., Sgt
AFHRL/SM, Lackland AFB, Texas 78236

GEORGE, Edward L.
USA Enlisted Evaluation Center, Ft Benjamin Harrison, Indianapolis, Indiana

GERMADNIK, George J., Sgt
AFHRL/PE, Lackland AFB, Texas 78236

GIORGIA, Joyce
AFHRL/OR, Lackland AFB, Texas 78236

GITHENS, Dr. William H.
U. S. Navy Personnel Research & Development Center, San Diego, California

GOES, Eugene H.
AFHRL/SM, Lackland AFB, Texas 78236

GOLDMAN, Dr. Lawrence A.
Naval Personnel Research & Development Laboratory, WNY, Washington, D. C. 20390

GORDON, M. M.
AFROTC/EDV (ACME), Maxwell AFB, Alabama 36112

GOTT, C. Deene
AFHRL/SM, Lackland AFB, Texas 78236

GOULD, R. Bruce
AFHRL/OR, Lackland AFB, Texas 78236

GRANT, Edward A., MSgt
AFHRL/SM, Lackland AFB, Texas 78236

GREENE, C. H., Lt JG
U. S. Coast Guard Training Center, Cape May, New Jersey 08204

GREENWAY, Charles A.
AFHRL/SM, Lackland AFB, Texas 78236

GROFF, Steven D., 1Lt
AFHRL/DOGT, Brooks AFB, Texas 78235

GUINN, Dr. Nancy
AFHRL/PE, Lackland AFB, Texas 78236

GUTCHESS, Robert, Lt Col
OTS/EV, Lackland AFB, Texas 78236

HAHN, Clifford P.
American Institutes for Research, 8555 16th Street, Silver Spring, Maryland 20910

HALTRECHT, Dr. E. J.
Pers Rsch Off, Rm GU-146, Ontario Hydro, 620 University Avenue, Toronto 2, Ontario, Canada
HANEY, Charles W., Lt Col
HQ AFROTC-ED, Maxwell AFB, Alabama 36112
HAWKINS, Michael D., Capt
AFHRL/MD, Lackland AFB, Texas 78236
HAYES, John R.
URS/Matrix Company, Systems Research Division, 7245 Arlington Boulevard, Falls Church, Virginia 22042
HAZEL, Dr. Joe T.
AFHRL/OR, Lackland AFB, Texas 78236
HEAD, Lawrence W., Jr.
USAF School of Applied Cryptologic Sciences (TTSM), Goodfellow AFB, Texas 76901
HINSON, Gerald C., Cdr
U. S. Coast Guard Training Center, Governors Island, New York, New York 10004
HOGGATT, Ralph S., Col
AFVTG, HQS ATC/RST, Randolph AFB, Texas 78148
HOLDEN, Dr. John J.
USA Ordnance Center & School, ATSL-DTP, Aberdeen Proving Ground, Maryland 21005
HOLIHEN, Joseph M., Lt Col
USA Enlisted Evaluation Center, Ft Benjamin Harrison, Indianapolis, Indiana
HORAN, William T., Lt JG
USCG HQ, Testing Section, 400 7th Street, S. W., Washington, D. C.
HOSKINS, James A., Capt
AFHRL/MD, Lackland AFB, Texas 78236
HOUTZ, John C.
Army Recruiting Command, USARCRD-M, Ft Sheridan, Illinois 60037
HUGHES, John, Amn
AFHRL/MD, Lackland AFB, Texas 78236
HURT, Larry B., Sgt
AFHRL/SM, Lackland AFB, Texas 78236
JACKSON, Dr. Charles W.
Educ. Advisor, ATSAR-EA, U. S. Army Armored School, Fort Knox, Kentucky 40121
JAIMES, Leonard N.
AFHRL/DOJL, Lackland AFB, Texas 78236
JENSEN, Harald E.
AFHRL/PE, Lackland AFB, Texas 78236
JOHNSON, Allan L., Sgt
AFHRL/PE, Lackland AFB, Texas 78236
JOHNSON, C. Amos
AFHRL/MD, Lackland AFB, Texas 78236
JOHNSON, Veskal E., Lt Col
AFHRL/MD, Lackland AFB, Texas 78236

JONES, Dr. Earl I.
Navy Personnel Research and Development Center, San Diego, California
KANNER, Dr. Joseph H.
TRADOC, ATTS-EA, Fort Monroe, Virginia 23651
KARIKER, D.
Interstate Electronics Corporation (ANA), Anaheim, California 92805
KEETH, James B.
3700 OCMST, Lackland AFB, Texas 78236
KELLEHER, Anna May
U. S. Coast Guard Training Center, Governors Island, New York, New York 10004
KELLOGG, Verna S., II, Maj
HQ ATC RST, Randolph AFB, Texas 78148
KERR, Dr. Norman J.
Naval Technical Training Research Branch, Naval Air Station, Millington, Tennessee 38054
KNIGHT, James R., Capt
USAF Academy/DFLS, Colorado Springs, Colorado 80840
KNOUSE, Stephen B., Capt
AFHRL/MD, Lackland AFB, Texas 78236
KNUTSON, Cecilia
AFHRL/SM, Lackland AFB, Texas 78236
KOPLYAY, Dr. Janos B.
AFHRL/SM, Lackland AFB, Texas 78236
KOYM, Kenneth K.
AFHRL/OR, Lackland AFB, Texas 78236
KREIGER, Gregory M.
USA Security Agency, Ft Devens, Massachusetts 01433
LANTERMAN, Richard
Office of Personnel, U. S. Coast Guard, Washington, D. C. 20590
LAWRENCE, Gary R., Capt
3700 OCMST, Lackland AFB, Texas 78236
LAWRENCE, Harry G.
3700 OCMSS, Lackland AFB, Texas 78236
LEATHERMAN, Dr. C. D.
Hq Defense Language Institute, WNY Anacostia Annex, Washington, D. C. 20374
LECZNAR, William B.
AFRHL/OR, Lackland AFB, Texas 78236
LEECH, E. L.
Naval Training Support (OIA), Bldg 45, Naval Air Station, Pensacola, Florida 32508
LEWIS, Nancy A.
3700 OCMST, Lackland AFB, Texas 78236
LILLIE, Richard A., Amn
AFHRL/SM, Lackland AFB, Texas 78236
LONGRIDGE, Thomas M., Jr., Capt
USAF School of Applied Aerospace Science, Training Research Applications Branch/TTOE/20, Sheppard AFB, Texas 76311
LOOPER, Larry T.
AFHRL/MD, Lackland AFB, Texas 78236

LOZANO, Fernando
AFHRL/DOJZ, Brooks AFB, Texas 78236
MACALUSO, Charles
Naval Examining Center, Bldg 2711, Great Lakes, Illinois 60088
MAGDARZ, E. F.
IBM Corporation, Department 829, Box 12195, Research Triangle Park, North Carolina 27709
MAJESTY, Melvin S., Lt Col
SMS-O, Lackland AFB, Texas 78236
MALDONADO, Diana L.
AFHRL/MD, Lackland AFB, Texas 78236
MALONE, Richard F., Maj
USA Institute of Admin, Eval Br, Bldg 400, Ft Benjamin Harrison, Indiana 46216
MARTINEZ, Ray
AFHRL/SM, Lackland AFB, Texas 78236
MASSEY, Iris
AFHRL/PE, Lackland AFB, Texas 78236
MATHEWS, Maj
AIAOS (Acad Instr & Allied Off School), Maxwell AFB, Alabama 36112
MATHON, William S., AIC
AFHRL/SM, Lackland AFB, Texas 78236
MAYO, Dr. George D.
Staff CNATECHTRA, NAS Memphis 75, Millington, Tennessee 38054
MCDONALD, Maj
AIAOS (Acad Instr & Allied Off School), Bldg 803, Maxwell AFB, Alabama 36112
MCDOWELL, Walter E.
ODCSTS, HQ TRADOC, Ft Monroe, Virginia, 864 Elder Road, Newport News, Virginia 23602
MCFARLAND, Barry P., 1Lt
AFHRL/OR, Lackland AFB, Texas 78236
MCLANATHAN, Frank L.
Department of Psychology, St Mary's University, San Antonio, Texas 78274
MCNATT, Donald R., SSgt
AFHRL/SM, Lackland AFB, Texas 78236
MCQUISTON, Fred C., Amn
AFHRL/SM, Lackland AFB, Texas 78236
MCREYNOLDS, Dr. Jane
AFHRL/DO, Brooks AFB, Texas 78236
MEAD, Don, Lt Col
HQ ATC/XPT, Randolph AFB, Texas 78148
MEEHAN, Dr. John
Air Univ - Evaluation & Research, AU/EDU, Maxwell AFB, Alabama 36112
MICHELSON, Alan E., Sgt
AFHRL/PE, Lackland AFB, Texas 78236
MILLER, James P., Cdr
Navy Occupational Task Analysis Program, Washington, D. C. 20370
MILLER, Joseph B., Sgt
AFHRL/SM, Lackland AFB, Texas 78236

MILLER, Dr. Robert E.
AFHRL/PE, Lackland AFB, Texas 78236

MOONAN, Dr. William J.
Personnel Systems Research Department, U. S. Naval Personnel Research Activity, San Diego, California 92152

MOSHER, Charles B., LCDR
U. S. Coast Guard Institute, P. O. Substation 18, Oklahoma City, Oklahoma 73169

MOUW, Robert L., Sgt
AFHRL/SM, Lackland AFB, Texas 78236

MULLER, Richard L.
Naval Education and Training, Code N-3522, NAS, Pensacola, Florida 32508

MULLINS, Dr. Cecil J.
AFHRL/PE, Lackland AFB, Texas 78236

MUNDY, Max
Interstate Electronics Corp (ETR), Anaheim, California 92805

MURPHY, Dr. Donald B.
System Development Corporation, 2500 Colorado Avenue, Santa Monica, California 90406

MURPHY, James H., Cdr
Naval Technical Training Evaluation Branch, Naval Air Station, Millington, Tennessee 38054

MYER, Jacob E.
AFHRL/SM, Lackland AFB, Texas 78236

NALL, Willian N., Maj
Tng Programs Br, Doc & Tng Dev Div, Combat & Tng Dev, USASSESS, USA Southeastern Signal School, Fort Gordon, Georgia 30905

NAURATH, David A.
Code 5342, Human Factors Eng Br, Lab Department, Naval Missile Center, Point Mugu, California 93042

NEWTON, Dr. Peter
NSA/Central Security Service, Ft George G. Meade, Maryland 20755

NOBLE, Bart E., 2Lt
AFHRL/PE, Lackland AFB, Texas 78236

NORDHAUSER, Fred, Maj
AFHRL/MD, Lackland AFB, Texas 78236

NORTHEY, W. H., Cdr
Director, General Applied Research (CEHQ), Ottawa, Ontario, Canada

O'BRIEN, Robert F.
Defense Language Institute, English Language Branch, Lackland AFB, Texas 78236

OETTING, George, Lt Col
Extension Course Institute (EDXV), Gunter AFB, Alabama 36114

OSBORN, William C.
HUMRRO Div 2, Fort Knox, Kentucky 40121

OTJEN, J. P., Capt
USA Infantry School, ATSH-D-ED, Fort Benning, Georgia 31905

PARKS, Cortez
Defense Language Institute, English Language Branch, Lackland AFB, Texas 78236

PARRIS, Dr. Howard L.
AFHRL/CCN, Brooks AFB, Texas 78235
PEARSALL, Glen, Yn
U. S. Coast Guard Training Center, Governors Island, New York, New York 10004
PEARSON, Hampton R., Lt Col
USA Mil Pers Ctr Mil Occ Bk, DAPC-PMR, 2461 Eisenhower Avenue, Alexandria, Virginia 22331
PETERSEN, Ronald C., Capt
U. S. Army Biomedical Lab, Edgewood Arsenal, Aberdeen Proving Ground, Maryland
PENNELL, Dr. Roger L.
AFHRL/TT, Lowry AFB, Colorado 80230
PHALEN, William J.
AFHRL/OR, Lackland AFB, Texas 78236
PINA, Manuel
AFHRL/MD, Lackland AFB, Texas 78236
PISTEL, Louis L.
AFHRL/DOJ, Brooks AFB, Texas 78235
PORTER, George V., Jr.
USAF Academy, Colorado Springs, Colorado 80840
POWELL, Harold B., Col
AFHRL/DO, Brooks AFB, Texas 78235
POWELL, Joe M., TSgt
AFHRL/SM, Lackland AFB, Texas 78236
POWELL, Patrick W., Sgt
AFHRL/SM, Lackland AFB, Texas 78236
PRIM, David F., Maj
Evaluations Branch, U. S. Army MP School, Fort Gordon, Georgia 30905
PURRY, B. N., Sqn Ldr
University of Edinburgh, Department of Politics, William Robertson Bldg, George Square, Edinburgh EH8 9jy, Great Britain
RAMBO, George M., Sgt
AFHRL/SM, Lackland AFB, Texas 78236
RATLIFF, Forrest R., Lt Col
AFHRL/MD, 701 Prince Street, Alexandria, Virginia 22314
REINBERG, David, Lt Col
AFHRL/PE, Lackland AFB, Texas 78236
RICK, Ronald A., Maj
OMU (AOIH) HQMC, Marine Corps Base, Quantico, Virginia 22134
RILEY, Eldon W.
Naval Education & Training NAS, Pensacola, Florida 32508
ROACH, James F., 1Lt
AFHRL/MD, Lackland AFB, Texas 78236
ROGERS, Charles R.
AFHRL/SM, Lackland AFB, Texas 78236
ROGERS, John R.
AFHRL/SM, Lackland AFB, Texas 78236

ROUGHGARDEN, Kenneth, Cdr
U. S. Coast Guard Institute, P. O. Substation 18, Oklahoma City, Oklahoma 73169

RUBANO, Louis, Lt
Service Sch Cmd, NTC, Naval Training Center, Orlando, Florida 32813

SAKO, Sydney
Defense Language Institute, English Language Branch, Lackland AFB, Texas 78236

SANDERS, Cynthia L., 2Lt
Community College of the Air Force, Randolph AFB, Texas 78148

SANDS, William A.
Navy Personnel Research & Development Center, San Diego, California 92152

SATHER, John, Sgt
AFHRL/SM, Lackland AFB, Texas 78236

SCHARF, George P.
Training Research Branch, USAF SSAS, Chanute AFB, Illinois 61866

SCRUGGS, James T., Maj
Pers Mgt Dev Dir USA Mil Pers Ctr, 2nd & R Street, S. W., Tempo A Building, Washington, D. C. 20315

SECRIST, G. E., Maj
AFHRL/PE, Lackland AFB, Texas 78236

SEIVERT, William D., Sgt
AFHRL/SM, Lackland AFB, Texas 78236

SEUBERLICH, H. E., Lt Col
Federal Ministry of Defence, 53 Bonn, Germany, Postfach 161

SHAFER, Joe
AFMTC/CDCCI, Lackland AFB, Texas 78236

SHORE, C. W., Maj
ATC/XPTT, Randolph AFB, Texas 78148

SHUFORD, Emir H., Jr.
RAND Corporation, 1700 Main Street, Santa Monica, California 90406

SIBLEY, W. L.
Info Sciences & Math Dept, RAND, Santa Monica, California 90406

SILVA, Elida A.
AFHRL/DOJL, Lackland AFB, Texas 78236

SLACK, Maj
Office of Manpower Utilization (A01H) (A Field Activity, G-1 Div) Quantico, Virginia 22134

SLEBODNICK, E. B.
Education & Training, Boeing Aerospace Company, P. O. Box 3999, Mail Stop 8K-01, Seattle, Washington 98124

SMITH, Dr. Kay H.
USA Research Institute, U. S. Army Infantry School, Fort Benning Field Unit, Fort Benning, Georgia 31905

SMITH, Mabron
ATSMM-DL, U. S. Army Missile & Munitions Center & School, Redstone Arsenal, Alabama 35809

SMITH, Presley E., AlC
AFHRL/SM, Lackland AFB, Texas 78236

SMITH, R. E.
U. S. Civil Service Commission, Dallas, Texas 75202

SMITH, Dr. Robert G.
HUMRRO, 300 N. Washington, Alexandria, Virginia 22314

SNELL, Larry L., Sgt
AFHRL/SM, Lackland AFB, Texas 78236

SNYDER, Quay C., Lt Col
Box 497, U. S. Army War College, Carlisle Barracks, Pennsylvania 17013

SOLOMON, William, Amn
AFHRL/OR, Lackland AFB, Texas 78236

SOUTER, Jimmy D.
AFHRL/SM, Lackland AFB, Texas 78236

STACEY, William J.
AFHRL/OR, Lackland AFB, Texas 78236

STAUFFER, Glenn F., Col
Community College of the Air Force, Randolph AFB, Texas 78148

STEFANCYK, Paul, Amn
AFHRL/SM, Lackland AFB, Texas 78236

STEPHENSON, Dr. Robert W.
6007 Bradley Boulevard, Bethesda, Maryland 20034

STIETZEL, Walter, CMSgt
AFHRL/SM, Lackland AFB, Texas 78236

STOCK, William E.
Program Planning & Development, V-T Ed, Dept of Education, Capitol Square, St Paul, Minnesota 55101

STOFFEL, Nancy
Minnesota Operation MEDIHC, 390 North Robert Street, St Paul, Minnesota

STRIGARI, Anthony J.
USA Enlisted Evaluation Center, Ft Benjamin Harrison, Indianapolis, Indiana

SWENEY, Dr. Arthur B.
Center of Human Appraisal, Wichita State University, Wichita, Kansas 67208

SWINEY, John F., Jr., Capt
USAFA, Department Life & Behavioral Science, P. O. Box 79, Colorado Springs, Colorado 80840

TAYLOR, John N.
AFHRL/MD, Lackland AFB, Texas 78236

TELISAK, Philip
AFHRL/LG, Brooks AFB, Texas 78235

TEMPLE, Narcie
AFHRL/PE, Lackland AFB, Texas 78236

THOMPSON, Ronnie A., Sgt
AFHRL/PE, Lackland AFB, Texas 78236

TICE, Thomas E., Capt
Sys Dev & Eval Office TAGCEN, Rm GB137, The Forrestal Bldg, HQDA, Washington, D. C. 20314

TIMMENS, James M., Lt Col
USA Personnel & Administration, Combat Developments Activity, Fort Benjamin Harrison, Indiana 46249

TINDALL, J. E., Jr.
Data-Design Laboratories, S Koger Exec Ctr, Norfolk, Virginia 23502

TITSWORTH, William
AFHRL/OR, Lackland AFB, Texas 78236

TOEWS, A. J.
OIC Manpower Planning, Royal Canadian Mounted Police, 250 Tremblay Road, Ottawa, Ontario, K1A OR2 Canada

TRAWEEK, John
OTS/EV, Lackland AFB, Texas 78236

TURGEON, Donald N., Sgt
AFHRL/SM, Lackland AFB, Texas 78236

TUTTLE, Dr. Thomas C.
Westinghouse Behavioral Safety Center, Box 948, American City Boulevard, Columbia, Maryland 21044

VALENTINE, Dr. Lonnie D.
AFHRL/PE, Lackland AFB, Texas 78236

VAN CLEVE, Roy R., Col
Rt 9, Box 7 37 E, Austin, Texas

VARNEY, Nicholas C., Capt
AFHRL/PE, Lackland AFB, Texas 78236

VEAUDRY, Wallace, Col
Assessment Center Project, U. S. Army Infantry School, Ft Benning, Georgia 31905

VINEBERG, Dr. Robert
HUMRRO P. O. Box 5787, Presidio of Monterey, California 93940

VINSON, Dr. David B.
1210 Medical Towers, Houston, Texas 77025

VITOLA, Bart M.
AFHRL/PE, Lackland AFB, Texas 78236

WAAG, Dr. Wayne L.
AFHRL/Flying Training Division, Williams AFB, Arizona

WADE, James E., Maj
AFHRL/DOGP, Brooks AFB, Texas 78235

WAGGENER, Dr. Robert R.
Counseling & Guidance, U. S. Armed Forces Institute, 2318 S. Park Street, Madison, Wisconsin 53713

WALDKOETTER, Raymond O.
7633 South Singleton Drive, Indianapolis, Indiana

WALKER, C. L., Jr., LCDR
Central Test Site for the Pers & Trng Eval Prog, Naval Guided Missiles School, Dam Neck, Virginia Beach, Virginia 23461

WALKER, Lewis A.
AFHRL/SM, Lackland AFB, Texas 78236
WALKER, Neilson P.
AFHRL/SM, Lackland AFB, Texas 78236
WANCE, William W.
USA Enlisted Evaluation Center, Ft Benjamin Harrison, Indiana 46249
WARD, Dr. Joe H., Jr.
AFHRL/MD, Lackland AFB, Texas 78236
WATSON, W. J., Sqn Ldr
Royal Australian AF, Occupational Research Division, AFHRL, Lackland AFB, Texas 78236
WEBB, C. K.
McDonnell Douglas Astronautics Co, St Louis, Missouri 63166
WEBBER, Robert E., Sgt
AFHRL/MD, Lackland AFB, Texas 78236
WEEKS, Joseph L., Sgt
AFHRL/PE, Lackland AFB, Texas 78236
WEISSMULLER, Johnny J., Sgt
AFHRL/SM, Lackland AFB, Texas 78236
WELLS, Harry E., Maj
OMU (AOIH) HQMC, Marine Corps Base, Quantico, Virginia 22134
WELLS, W. L.
AFHRL/SM, Lackland AFB, Texas 78236
WEST, Charles M.
USA PACDA-HRD, Fort Harrison, Indiana 46249
WETHY, Roy B.
USA Mil Pers Ctr - Mil Occ Bk, DAPC-PMR, 2461 Eisenhower Avenue, Alexandria, Virginia 22331
WHITE, Donald T., MSgt
USA Medical Equipment & Optical School, Denver, Colorado 80240
WHITEHEAD, Larry K.
AFHRL/SM, Lackland AFB, Texas 78236
WILBOURN, James M.
AFHRL/PE, Lackland AFB, Texas 78236
WILEY, Llewellyn N.
AFHRL/OR, Lackland AFB, Texas 78236
WILFONG, Dr. Harry D.
HQ ATC RST, Randolph AFB, Texas 78148
WILLIAMS, J. E. C., LCDR
Staff Australian Naval Attache of Australian Embassy, 1601 Massachusetts Avenue, Washington, D. C. 20036
WILLIAMS, Rayburn A.
Naval Training, RDT&E Div, Bldg #624, NAS, Pensacola, Florida 32508
WILLING, Dr. Richard
U. S. Coast Guard Institute, P. O. Substation 18, Oklahoma City, Oklahoma 73169
WOOD, G. David, SSgt
AFHRL/OR, Lackland AFB, Texas 78236

WOOLDRIDGE, Larry A., Capt
OMU (AOIH) HQMC, Marine Corps Base, Quantico, Virginia 22134
ZACHERT, Virginia
Route 1, Box 28, Good Hope, Georgia 30641
ZUMBRO, Patrick, Amn
AFHRL/OR, Lackland AFB, Texas 78236
ZURO, Joseph E., Col
Air War College, Air University, Box 309, Maxwell AFB, Alabama 36113